## **PCT**

# WORLD INTELLECTUAL PROPERTY ORGANIZATION International Bureau



# INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6:

C12N 15/31, C07K 14/22, 16/12, A61K 39/095 // (C12N 15/31, C12R 1:36)

(11) International Publication Number:

WO 99/24578

**A2** 

(43) International Publication Date:

20 May 1999 (20.05.99)

(21) International Application Number:

PCT/IB98/01665

(22) International Filing Date:

9 October 1998 (09.10.98)

(30) Priority Data:

tionity Data.				
9723516.2	6 November 1997 (06.11.97)	GB		
9724190.5	14 November 1997 (14.11.97)	GB		
9724386.9	724386.9 18 November 1997 (18.11.97)			
9725158.1		GB		
9726147.3	10 December 1997 (10.12.97)	GB		
9800759.4	14 January 1998 (14.01.98)	GB		
9819016.8	1 September 1998 (01.09.98)	GB		
	•			

(71) Applicant (for all designated States except US): CHIRON S.P.A. [IT/IT]; Via Fiorentina, 1, I-53100 Siena (IT).

(72) Inventors; and

(75) Inventors/Applicants (for US only): MASIGNANI, Vega [IT/IT]; Via Pantaneto, 105, I-53100 Siena (IT). RAP-PUOLI, Rino [IT/IT]; Via delle Rocche, 1, Vagliagli, I-53019 Castelnuovo Berardenga (IT). PIZZA, Mariagrazia [IT/IT]; Strada di Montalbuccio, 160, I-53100 Siena (IT). SCARLATO, Vincenzo [IT/IT]; Via Firenze, 3/37, I-53134 Colle Val d'Elsa (IT). GRANDI, Guido [IT/IT]; 9° Strada, 4, I-20090 Segrate (IT).

(74) Agent: HALLYBONE, Huw, George; Carpmaels & Ransford, 43 Bloomsbury Square, London WC1A 2RA (GB).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

#### **Published**

Without international search report and to be republished upon receipt of that report.

(54) Title: NEISSERIAL ANTIGENS

(57) Abstract

The invention provides proteins from *Neisseria meningitidis* (strains A and B) and from *Neisseria gonorrhoerae* including amino acid sequences, the corresponding nucleotide sequences, expression data, and serological data. The proteins are useful antigens for vaccines, immunogenic compositions, and/or diagnostics.

# FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

4.7	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AL AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
ΑŬ	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav	TM	Turkmenistan
BF	Burkina Faso	GR	Greece		Republic of Macedonia	TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	zw	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's	NZ	New Zealand		
CM	Cameroon		Republic of Korea	PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
cz	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		
EE	Latonia						

### **NEISSERIAL ANTIGENS**

This invention relates to antigens from Neisseria bacteria.

#### **BACKGROUND ART**

5

10

Neisseria meningitidis and Neisseria gonorrhoeae are non-motile, gram negative diplococci that are pathogenic in humans. N.meningitidis colonises the pharynx and causes meningitis (and, occasionally, septicaemia in the absence of meningitis); N.gonorrhoeae colonises the genital tract and causes gonorrhea. Although colonising different areas of the body and causing completely different diseases, the two pathogens are closely related, although one feature that clearly differentiates meningococcus from gonococcus is the presence of a polysaccharide capsule that is present in all pathogenic meningococci.

N.gonorrhoeae caused approximately 800,000 cases per year during the period 1983-1990 in the United States alone (chapter by Meitzner & Cohen, "Vaccines Against Gonococcal Infection", In: New Generation Vaccines, 2nd edition, ed. Levine, Woodrow, Kaper, & Cobon, Marcel Dekker, New York, 1997, pp.817-842). The disease causes significant morbidity but limited mortality.
Vaccination against N.gonorrhoeae would be highly desirable, but repeated attempts have failed. The main candidate antigens for this vaccine are surface-exposed proteins such as pili, porins, opacity-associated proteins (Opas) and other surface-exposed proteins such as the Lip, Laz, IgA1 protease and transferrin-binding proteins. The lipooligosaccharide (LOS) has also been suggested as vaccine (Meitzner & Cohen, supra).

N.meningitidis causes both endemic and epidemic disease. In the United States the attack rate is 0.6-1 per 100,000 persons per year, and it can be much greater during outbreaks (see Lieberman et al. (1996) Safety and Immunogenicity of a Serogroups A/C Neisseria meningitidis Oligosaccharide-Protein Conjugate Vaccine in Young Children. JAMA 275(19):1499-1503; Schuchat et al (1997) Bacterial Meningitis in the United States in 1995. N Engl J Med 337(14):970-976). In developing countries, endemic disease rates are much higher and during epidemics incidence rates can reach 500 cases per 100,000 persons per year. Mortality is extremely high, at 10-20% in the United States, and much higher in developing countries. Following the introduction of the conjugate vaccine against Haemophilus influenzae, N. meningitidis is the major cause of bacterial meningitis at all ages in the United States (Schuchat et al (1997) supra).

Based on the organism's capsular polysaccharide, 12 serogroups of *N.meningitidis* have been identified. Group A is the pathogen most often implicated in epidemic disease in sub-Saharan Africa. Serogroups B and C are responsible for the vast majority of cases in the United States and in most developed countries. Serogroups W135 and Y are responsible for the rest of the cases in the United States and developed countries. The meningococcal vaccine currently in use is a tetravalent polysaccharide vaccine composed of serogroups A, C, Y and W135. Although efficacious in adolescents and adults, it induces a poor immune response and short duration of protection, and cannot be used in infants [eg. Morbidity and Mortality weekly report, Vol.46, No. RR-5 (1997)]. This is because polysaccharides are T-cell independent antigens that induce a weak immune response that cannot be boosted by repeated immunization. Following the success of the vaccination against *H.influenzae*, conjugate vaccines against serogroups A and C have been developed and are at the final stage of clinical testing (Zollinger WD "New and Improved Vaccines Against Meningococcal Disease" in: New Generation Vaccines, supra, pp. 469-488; Lieberman et al (1996) supra; Costantino et al (1992) Development and phase I clinical testing of a conjugate vaccine against meningococcus A and C. Vaccine 10:691-698).

5

10

15

20

25

30

Meningococcus B remains a problem, however. This serotype currently is responsible for approximately 50% of total meningitis in the United States, Europe, and South America. The polysaccharide approach cannot be used because the menB capsular polysaccharide is a polymer of α(2-8)-linked N-acetyl neuraminic acid that is also present in mammalian tissue. This results in tolerance to the antigen; indeed, if an immune response were elicited, it would be anti-self, and therefore undesirable. In order to avoid induction of autoimmunity and to induce a protective immune response, the capsular polysaccharide has, for instance, been chemically modified substituting the N-acetyl groups with N-propionyl groups, leaving the specific antigenicity unaltered (Romero & Outschoorn (1994) Current status of Meningococcal group B vaccine candidates: capsular or non-capsular? Clin Microbiol Rev 7(4):559-575).

Alternative approaches to menB vaccines have used complex mixtures of outer membrane proteins (OMPs), containing either the OMPs alone, or OMPs enriched in porins, or deleted of the class 4 OMPs that are believed to induce antibodies that block bactericidal activity. This approach produces vaccines that are not well characterized. They are able to protect against the homologous strain, but are not effective at large where there are many antigenic variants of the outer membrane proteins. To overcome the antigenic variability, multivalent vaccines containing up to nine different

porins have been constructed (eg. Poolman JT (1992) Development of a meningococcal vaccine. Infect. Agents Dis. 4:13-28). Additional proteins to be used in outer membrane vaccines have been the opa and opc proteins, but none of these approaches have been able to overcome the antigenic variability (eg. Ala'Aldeen & Borriello (1996) The meningococcal transferrin-binding proteins 1 and 2 are both surface exposed and generate bactericidal antibodies capable of killing homologous and heterologous strains. Vaccine 14(1):49-53).

A certain amount of sequence data is available for meningococcal and gonoccocal genes and proteins (eg. EP-A-0467714, WO96/29412), but this is by no means complete. The provision of further sequences could provide an opportunity to identify secreted or surface-exposed proteins that are presumed targets for the immune system and which are not antigenically variable. For instance, some of the identified proteins could be components of efficacious vaccines against meningococcus B, some could be components of vaccines against all meningococcal serotypes, and others could be components of vaccines against all pathogenic Neisseriae.

#### THE INVENTION

5

10

20

15 The invention provides proteins comprising the Neisserial amino acid sequences disclosed in the examples. These sequences relate to *N.meningitidis* or *N.gonorrhoeae*.

It also provides proteins comprising sequences homologous (ie. having sequence identity) to the Neisserial amino acid sequences disclosed in the examples. Depending on the particular sequence, the degree of identity is preferably greater than 50% (eg. 65%, 80%, 90%, or more). These homologous proteins include mutants and allelic variants of the sequences disclosed in the examples. Typically, 50% identity or more between two proteins is considered to be an indication of functional equivalence. Identity between the proteins is preferably determined by the Smith-Waterman homology search algorithm as implemented in the MPSRCH program (Oxford Molecular), using an affine gap search with parameters gap open penalty=12 and gap extension penalty=1.

25 The invention further provides proteins comprising fragments of the Neisserial amino acid sequences disclosed in the examples. The fragments should comprise at least n consecutive amino acids from the sequences and, depending on the particular sequence, n is 7 or more (eg. 8, 10, 12, 14, 16, 18, 20 or more). Preferably the fragments comprise an epitope from the sequence.

The proteins of the invention can, of course, be prepared by various means (eg. recombinant expression, purification from cell culture, chemical synthesis etc.) and in various forms (eg. native, fusions etc.). They are preferably prepared in substantially pure or isolated form (ie. substantially free from other Neisserial or host cell proteins)

According to a further aspect, the invention provides antibodies which bind to these proteins. These may be polyclonal or monoclonal and may be produced by any suitable means.

According to a further aspect, the invention provides nucleic acid comprising the Neisserial nucleotide sequences disclosed in the examples. In addition, the invention provides nucleic acid comprising sequences homologous (*ie.* having sequence identity) to the Neisserial nucleotide sequences disclosed in the examples.

10

Furthermore, the invention provides nucleic acid which can hybridise to the Neisserial nucleic acid disclosed in the examples, preferably under "high stringency" conditions (eg. 65°C in a 0.1xSSC, 0.5% SDS solution).

Nucleic acid comprising fragments of these sequences are also provided. These should comprise at least n consecutive nucleotides from the Neisserial sequences and, depending on the particular sequence, n is 10 or more (eg 12, 14, 15, 18, 20, 25, 30, 35, 40 or more).

According to a further aspect, the invention provides nucleic acid encoding the proteins and protein fragments of the invention.

It should also be appreciated that the invention provides nucleic acid comprising sequences complementary to those described above (eg. for antisense or probing purposes).

Nucleic acid according to the invention can, of course, be prepared in many ways (eg. by chemical synthesis, from genomic or cDNA libraries, from the organism itself etc.) and can take various forms (eg. single stranded, double stranded, vectors, probes etc.).

In addition, the term "nucleic acid" includes DNA and RNA, and also their analogues, such as those containing modified backbones, and also peptide nucleic acids (PNA) etc.

5

10

According to a further aspect, the invention provides vectors comprising nucleotide sequences of the invention (eg. expression vectors) and host cells transformed with such vectors.

According to a further aspect, the invention provides compositions comprising protein, antibody, and/or nucleic acid according to the invention. These compositions may be suitable as vaccines, for instance, or as diagnostic reagents, or as immunogenic compositions.

The invention also provides nucleic acid, protein, or antibody according to the invention for use as medicaments (eg. as vaccines) or as diagnostic reagents. It also provides the use of nucleic acid, protein, or antibody according to the invention in the manufacture of: (i) a medicament for treating or preventing infection due to Neisserial bacteria; (ii) a diagnostic reagent for detecting the presence of Neisserial bacteria or of antibodies raised against Neisserial bacteria; and/or (iii) a reagent which can raise antibodies against Neisserial bacteria. Said Neisserial bacteria may be any species or strain (such as N. gonorrhoeae, or any strain of N. meningitidis, such as strain A, strain B or strain C).

The invention also provides a method of treating a patient, comprising administering to the patient a therapeutically effective amount of nucleic acid, protein, and/or antibody according to the invention.

According to further aspects, the invention provides various processes.

A process for producing proteins of the invention is provided, comprising the step of culturing a host cell according to the invention under conditions which induce protein expression.

A process for producing protein or nucleic acid of the invention is provided, wherein the the protein or nucleic acid is synthesised in part or in whole using chemical means.

A process for detecting polynucleotides of the invention is provided, comprising the steps of: (a) contacting a nucleic probe according to the invention with a biological sample under hybridizing conditions to form duplexes; and (b) detecting said duplexes.

A process for detecting proteins of the invention is provided, comprising the steps of: (a) contacting an antibody according to the invention with a biological sample under conditions suitable for the formation of an antibody-antigen complexes; and (b) detecting said complexes.

WO 99/24578 PCT/IB98/01665 -6-

A summary of standard techniques and procedures which may be employed in order to perform the invention (eg. to utilise the disclosed sequences for vaccination or diagnostic purposes) follows. This summary is not a limitation on the invention but, rather, gives examples that may be used, but are not required.

#### General 5

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of molecular biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature eg. Sambrook Molecular Cloning; A Laboratory Manual, Second Edition (1989); DNA Cloning, Volumes I and ii (D.N Glover ed. 1985); Oligonucleotide Synthesis (M.J. Gait ed, 1984); Nucleic Acid 10 Hybridization (B.D. Hames & S.J. Higgins eds. 1984); Transcription and Translation (B.D. Hames & S.J. Higgins eds. 1984); Animal Cell Culture (R.I. Freshney ed. 1986); Immobilized Cells and Enzymes (IRL Press, 1986); B. Perbal, A Practical Guide to Molecular Cloning (1984); the Methods in Enzymology series (Academic Press, Inc.), especially volumes 154 & 155; Gene 15 Transfer Vectors for Mammalian Cells (J.H. Miller and M.P. Calos eds. 1987, Cold Spring Harbor Laboratory); Mayer and Walker, eds. (1987), Immunochemical Methods in Cell and Molecular Biology (Academic Press, London); Scopes, (1987) Protein Purification: Principles and Practice, Second Edition (Springer-Verlag, N.Y.), and Handbook of Experimental Immunology, Volumes I-IV (D.M. Weir and C. C. Blackwell eds 1986).

Standard abbreviations for nucleotides and amino acids are used in this specification. 20

All publications, patents, and patent applications cited herein are incorporated in full by reference. In particular, the contents of UK patent applications 9723516.2, 9724190.5, 9724386.9, 9725158.1, 9726147.3, 9800759.4, and 9819016.8 are incorporated herein.

#### **Definitions**

A composition containing X is "substantially free of" Y when at least 85% by weight of the total 25 X+Y in the composition is X. Preferably, X comprises at least about 90% by weight of the total of X+Y in the composition, more preferably at least about 95% or even 99% by weight.

The term "comprising" means "including" as well as "consisting" eg. a composition "comprising" X may consist exclusively of X or may include something additional to X, such as X+Y.

The term "heterologous" refers to two biological components that are not found together in nature. The components may be host cells, genes, or regulatory regions, such as promoters. Although the heterologous components are not found together in nature, they can function together, as when a promoter heterologous to a gene is operably linked to the gene. Another example is where a Neisserial sequence is heterologous to a mouse host cell. A further examples would be two epitopes from the same or different proteins which have been assembled in a single protein in an arrangement not found in nature.

An "origin of replication" is a polynucleotide sequence that initiates and regulates replication of polynucleotides, such as an expression vector. The origin of replication behaves as an autonomous unit of polynucleotide replication within a cell, capable of replication under its own control. An origin of replication may be needed for a vector to replicate in a particular host cell. With certain origins of replication, an expression vector can be reproduced at a high copy number in the presence of the appropriate proteins within the cell. Examples of origins are the autonomously replicating sequences, which are effective in yeast; and the viral T-antigen, effective in COS-7 cells.

A "mutant" sequence is defined as DNA, RNA or amino acid sequence differing from but having sequence identity with the native or disclosed sequence. Depending on the particular sequence, the degree of sequence identity between the native or disclosed sequence and the mutant sequence is preferably greater than 50% (eg. 60%, 70%, 80%, 90%, 95%, 99% or more, calculated using the Smith-Waterman algorithm as described above). As used herein, an "allelic variant" of a nucleic acid molecule, or region, for which nucleic acid sequence is provided herein is a nucleic acid molecule, or region, that occurs essentially at the same locus in the genome of another or second isolate, and that, due to natural variation caused by, for example, mutation or recombination, has a similar but not identical nucleic acid sequence. A coding region allelic variant typically encodes a protein having similar activity to that of the protein encoded by the gene to which it is being compared. An allelic variant can also comprise an alteration in the 5' or 3' untranslated regions of the gene, such as in regulatory control regions (eg. see US patent 5,753,235).

#### Expression systems

5

10

15

20

25

30

The Neisserial nucleotide sequences can be expressed in a variety of different expression systems; for example those used with mammalian cells, baculoviruses, plants, bacteria, and yeast.

#### i. Mammalian Systems

5

10

15

Mammalian expression systems are known in the art. A mammalian promoter is any DNA sequence capable of binding mammalian RNA polymerase and initiating the downstream (3') transcription of a coding sequence (eg. structural gene) into mRNA. A promoter will have a transcription initiating region, which is usually placed proximal to the 5' end of the coding sequence, and a TATA box, usually located 25-30 base pairs (bp) upstream of the transcription initiation site. The TATA box is thought to direct RNA polymerase II to begin RNA synthesis at the correct site. A mammalian promoter will also contain an upstream promoter element, usually located within 100 to 200 bp upstream of the TATA box. An upstream promoter element determines the rate at which transcription is initiated and can act in either orientation [Sambrook et al. (1989) "Expression of Cloned Genes in Mammalian Cells." In Molecular Cloning: A Laboratory Manual, 2nd ed.].

Mammalian viral genes are often highly expressed and have a broad host range; therefore sequences encoding mammalian viral genes provide particularly useful promoter sequences. Examples include the SV40 early promoter, mouse mammary tumor virus LTR promoter, adenovirus major late promoter (Ad MLP), and herpes simplex virus promoter. In addition, sequences derived from non-viral genes, such as the murine metallotheionein gene, also provide useful promoter sequences. Expression may be either constitutive or regulated (inducible), depending on the promoter can be induced with glucocorticoid in hormone-responsive cells.

The presence of an enhancer element (enhancer), combined with the promoter elements described above, will usually increase expression levels. An enhancer is a regulatory DNA sequence that can stimulate transcription up to 1000-fold when linked to homologous or heterologous promoters, with synthesis beginning at the normal RNA start site. Enhancers are also active when they are placed upstream or downstream from the transcription initiation site, in either normal or flipped orientation, or at a distance of more than 1000 nucleotides from the promoter [Maniatis et al. (1987) Science 236:1237; Alberts et al. (1989) Molecular Biology of the Cell, 2nd ed.]. Enhancer elements derived from viruses may be particularly useful, because they usually have a broader host range. Examples include the SV40 early gene enhancer [Dijkema et al (1985) EMBO J. 4:761] and the enhancer/promoters derived from the long terminal repeat (LTR) of the Rous Sarcoma Virus [Gorman et al. (1982b) Proc. Natl. Acad. Sci. 79:6777] and from human cytomegalovirus [Boshart et al. (1985) Cell 41:521]. Additionally, some enhancers are regulatable and become active only

5

10

in the presence of an inducer, such as a hormone or metal ion [Sassone-Corsi and Borelli (1986) *Trends Genet. 2*:215; Maniatis et al. (1987) *Science* 236:1237].

A DNA molecule may be expressed intracellularly in mammalian cells. A promoter sequence may be directly linked with the DNA molecule, in which case the first amino acid at the N-terminus of the recombinant protein will always be a methionine, which is encoded by the ATG start codon. If desired, the N-terminus may be cleaved from the protein by *in vitro* incubation with cyanogen bromide.

Alternatively, foreign proteins can also be secreted from the cell into the growth media by creating chimeric DNA molecules that encode a fusion protein comprised of a leader sequence fragment that provides for secretion of the foreign protein in mammalian cells. Preferably, there are processing sites encoded between the leader fragment and the foreign gene that can be cleaved either *in vivo* or *in vitro*. The leader sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell. The adenovirus triparite leader is an example of a leader sequence that provides for secretion of a foreign protein in mammalian cells.

- Usually, transcription termination and polyadenylation sequences recognized by mammalian cells are regulatory regions located 3' to the translation stop codon and thus, together with the promoter elements, flank the coding sequence. The 3' terminus of the mature mRNA is formed by site-specific post-transcriptional cleavage and polyadenylation [Birnstiel et al. (1985) Cell 41:349; Proudfoot and Whitelaw (1988) "Termination and 3' end processing of eukaryotic RNA. In
   Transcription and splicing (ed. B.D. Hames and D.M. Glover); Proudfoot (1989) Trends Biochem. Sci. 14:105]. These sequences direct the transcription of an mRNA which can be translated into the polypeptide encoded by the DNA. Examples of transcription terminater/polyadenylation signals include those derived from SV40 [Sambrook et al (1989) "Expression of cloned genes in cultured mammalian cells." In Molecular Cloning: A Laboratory Manual].
- Usually, the above described components, comprising a promoter, polyadenylation signal, and transcription termination sequence are put together into expression constructs. Enhancers, introns with functional splice donor and acceptor sites, and leader sequences may also be included in an expression construct, if desired. Expression constructs are often maintained in a replicon, such as an extrachromosomal element (eg. plasmids) capable of stable maintenance in a host, such as mammalian cells or bacteria. Mammalian replication systems include those derived from animal

viruses, which require trans-acting factors to replicate. For example, plasmids containing the replication systems of papovaviruses, such as SV40 [Gluzman (1981) Cell 23:175] or polyomavirus, replicate to extremely high copy number in the presence of the appropriate viral T antigen. Additional examples of mammalian replicons include those derived from bovine papillomavirus and Epstein-Barr virus. Additionally, the replicon may have two replicaton systems, thus allowing it to be maintained, for example, in mammalian cells for expression and in a prokaryotic host for cloning and amplification. Examples of such mammalian-bacteria shuttle vectors include pMT2 [Kaufman et al. (1989) Mol. Cell. Biol. 9:946] and pHEBO [Shimizu et al. (1986) Mol. Cell. Biol. 6:1074].

The transformation procedure used depends upon the host to be transformed. Methods for introduction of heterologous polynucleotides into mammalian cells are known in the art and include dextran-mediated transfection, calcium phosphate precipitation, polybrene mediated transfection, protoplast fusion, electroporation, encapsulation of the polynucleotide(s) in liposomes, and direct microinjection of the DNA into nuclei.

15 Mammalian cell lines available as hosts for expression are known in the art and include many immortalized cell lines available from the American Type Culture Collection (ATCC), including but not limited to, Chinese hamster ovary (CHO) cells, HeLa cells, baby hamster kidney (BHK) cells, monkey kidney cells (COS), human hepatocellular carcinoma cells (eg. Hep G2), and a number of other cell lines.

#### 20 ii. Baculovirus Systems

25

30

5

The polynucleotide encoding the protein can also be inserted into a suitable insect expression vector, and is operably linked to the control elements within that vector. Vector construction employs techniques which are known in the art. Generally, the components of the expression system include a transfer vector, usually a bacterial plasmid, which contains both a fragment of the baculovirus genome, and a convenient restriction site for insertion of the heterologous gene or genes to be expressed; a wild type baculovirus with a sequence homologous to the baculovirus-specific fragment in the transfer vector (this allows for the homologous recombination of the heterologous gene in to the baculovirus genome); and appropriate insect host cells and growth media.

After inserting the DNA sequence encoding the protein into the transfer vector, the vector and the wild type viral genome are transfected into an insect host cell where the vector and viral genome are allowed to recombine. The packaged recombinant virus is expressed and recombinant plaques

5

10

are identified and purified. Materials and methods for baculovirus/insect cell expression systems are commercially available in kit form from, *inter alia*, Invitrogen, San Diego CA ("MaxBac" kit). These techniques are generally known to those skilled in the art and fully described in Summers and Smith, *Texas Agricultural Experiment Station Bulletin No. 1555* (1987) (hereinafter "Summers and Smith").

Prior to inserting the DNA sequence encoding the protein into the baculovirus genome, the above described components, comprising a promoter, leader (if desired), coding sequence of interest, and transcription termination sequence, are usually assembled into an intermediate transplacement construct (transfer vector). This construct may contain a single gene and operably linked regulatory elements; multiple genes, each with its owned set of operably linked regulatory elements; or multiple genes, regulated by the same set of regulatory elements. Intermediate transplacement constructs are often maintained in a replicon, such as an extrachromosomal element (eg. plasmids) capable of stable maintenance in a host, such as a bacterium. The replicon will have a replication system, thus allowing it to be maintained in a suitable host for cloning and amplification.

- 15 Currently, the most commonly used transfer vector for introducing foreign genes into AcNPV is pAc373. Many other vectors, known to those of skill in the art, have also been designed. These include, for example, pVL985 (which alters the polyhedrin start codon from ATG to ATT, and which introduces a BamHI cloning site 32 basepairs downstream from the ATT; see Luckow and Summers, *Virology* (1989) 17:31.
- The plasmid usually also contains the polyhedrin polyadenylation signal (Miller et al. (1988) Ann. Rev. Microbiol., 42:177) and a prokaryotic ampicillin-resistance (amp) gene and origin of replication for selection and propagation in E. coli.

Baculovirus transfer vectors usually contain a baculovirus promoter. A baculovirus promoter is any DNA sequence capable of binding a baculovirus RNA polymerase and initiating the downstream (5' to 3') transcription of a coding sequence (eg. structural gene) into mRNA. A promoter will have a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region usually includes an RNA polymerase binding site and a transcription initiation site. A baculovirus transfer vector may also have a second domain called an enhancer, which, if present, is usually distal to the structural gene. Expression may be either regulated or constitutive.

5

10

15

20

25

30

Structural genes, abundantly transcribed at late times in a viral infection cycle, provide particularly useful promoter sequences. Examples include sequences derived from the gene encoding the viral polyhedron protein, Friesen et al., (1986) "The Regulation of Baculovirus Gene Expression," in: *The Molecular Biology of Baculoviruses* (ed. Walter Doerfler); EPO Publ. Nos. 127 839 and 155 476; and the gene encoding the p10 protein, Vlak et al., (1988), *J. Gen. Virol.* 69:765.

DNA encoding suitable signal sequences can be derived from genes for secreted insect or baculovirus proteins, such as the baculovirus polyhedrin gene (Carbonell et al. (1988) *Gene*, 73:409). Alternatively, since the signals for mammalian cell posttranslational modifications (such as signal peptide cleavage, proteolytic cleavage, and phosphorylation) appear to be recognized by insect cells, and the signals required for secretion and nuclear accumulation also appear to be conserved between the invertebrate cells and vertebrate cells, leaders of non-insect origin, such as those derived from genes encoding human α-interferon, Maeda et al., (1985), *Nature 315*:592; human gastrin-releasing peptide, Lebacq-Verheyden et al., (1988), *Molec. Cell. Biol. 8*:3129; human IL-2, Smith et al., (1985) *Proc. Nat'l Acad. Sci. USA*, 82:8404; mouse IL-3, (Miyajima et al., (1987) *Gene 58*:273; and human glucocerebrosidase, Martin et al. (1988) *DNA*, 7:99, can also be used to provide for secretion in insects.

A recombinant polypeptide or polyprotein may be expressed intracellularly or, if it is expressed with the proper regulatory sequences, it can be secreted. Good intracellular expression of nonfused foreign proteins usually requires heterologous genes that ideally have a short leader sequence containing suitable translation initiation signals preceding an ATG start signal. If desired, methionine at the N-terminus may be cleaved from the mature protein by *in vitro* incubation with cyanogen bromide.

Alternatively, recombinant polyproteins or proteins which are not naturally secreted can be secreted from the insect cell by creating chimeric DNA molecules that encode a fusion protein comprised of a leader sequence fragment that provides for secretion of the foreign protein in insects. The leader sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the translocation of the protein into the endoplasmic reticulum.

After insertion of the DNA sequence and/or the gene encoding the expression product precursor of the protein, an insect cell host is co-transformed with the heterologous DNA of the transfer vector and the genomic DNA of wild type baculovirus -- usually by co-transfection. The promoter

and transcription termination sequence of the construct will usually comprise a 2-5kb section of the baculovirus genome. Methods for introducing heterologous DNA into the desired site in the baculovirus virus are known in the art. (See Summers and Smith *supra*; Ju et al. (1987); Smith et al., *Mol. Cell. Biol.* (1983) 3:2156; and Luckow and Summers (1989)). For example, the insertion can be into a gene such as the polyhedrin gene, by homologous double crossover recombination; insertion can also be into a restriction enzyme site engineered into the desired baculovirus gene. Miller et al., (1989), *Bioessays 4*:91. The DNA sequence, when cloned in place of the polyhedrin gene in the expression vector, is flanked both 5' and 3' by polyhedrin-specific sequences and is positioned downstream of the polyhedrin promoter.

5

- The newly formed baculovirus expression vector is subsequently packaged into an infectious 10 recombinant baculovirus. Homologous recombination occurs at low frequency (between about 1% and about 5%); thus, the majority of the virus produced after cotransfection is still wild-type virus. Therefore, a method is necessary to identify recombinant viruses. An advantage of the expression system is a visual screen allowing recombinant viruses to be distinguished. The polyhedrin protein, which is produced by the native virus, is produced at very high levels in the nuclei of infected cells 15 at late times after viral infection. Accumulated polyhedrin protein forms occlusion bodies that also contain embedded particles. These occlusion bodies, up to 15 µm in size, are highly refractile, giving them a bright shiny appearance that is readily visualized under the light microscope. Cells infected with recombinant viruses lack occlusion bodies. To distinguish recombinant virus from 20 wild-type virus, the transfection supernatant is plaqued onto a monolayer of insect cells by techniques known to those skilled in the art. Namely, the plaques are screened under the light microscope for the presence (indicative of wild-type virus) or absence (indicative of recombinant virus) of occlusion bodies. "Current Protocols in Microbiology" Vol. 2 (Ausubel et al. eds) at 16.8 (Supp. 10, 1990); Summers and Smith, supra; Miller et al. (1989).
- Recombinant baculovirus expression vectors have been developed for infection into several insect cells. For example, recombinant baculoviruses have been developed for, inter alia: Aedes aegypti , Autographa californica, Bombyx mori, Drosophila melanogaster, Spodoptera frugiperda, and Trichoplusia ni (WO 89/046699; Carbonell et al., (1985) J. Virol. 56:153; Wright (1986) Nature 321:718; Smith et al., (1983) Mol. Cell. Biol. 3:2156; and see generally, Fraser, et al. (1989) In Vitro Cell. Dev. Biol. 25:225).

Cells and cell culture media are commercially available for both direct and fusion expression of heterologous polypeptides in a baculovirus/expression system; cell culture technology is generally known to those skilled in the art. See, eg. Summers and Smith supra.

The modified insect cells may then be grown in an appropriate nutrient medium, which allows for stable maintenance of the plasmid(s) present in the modified insect host. Where the expression product gene is under inducible control, the host may be grown to high density, and expression induced. Alternatively, where expression is constitutive, the product will be continuously expressed into the medium and the nutrient medium must be continuously circulated, while removing the product of interest and augmenting depleted nutrients. The product may be purified by such techniques as chromatography, eg. HPLC, affinity chromatography, ion exchange chromatography, etc.; electrophoresis; density gradient centrifugation; solvent extraction, or the like. As appropriate, the product may be further purified, as required, so as to remove substantially any insect proteins which are also secreted in the medium or result from lysis of insect cells, so as to provide a product which is at least substantially free of host debris, eg. proteins, lipids and polysaccharides.

In order to obtain protein expression, recombinant host cells derived from the transformants are incubated under conditions which allow expression of the recombinant protein encoding sequence. These conditions will vary, dependent upon the host cell selected. However, the conditions are readily ascertainable to those of ordinary skill in the art, based upon what is known in the art.

#### iii. Plant Systems

5

10

There are many plant cell culture and whole plant genetic expression systems known in the art. Exemplary plant cellular genetic expression systems include those described in patents, such as: US 5,693,506; US 5,659,122; and US 5,608,143. Additional examples of genetic expression in plant cell culture has been described by Zenk, Phytochemistry 30:3861-3863 (1991). Descriptions of plant protein signal peptides may be found in addition to the references described above in Vaulcombe et al., Mol. Gen. Genet. 209:33-40 (1987); Chandler et al., Plant Molecular Biology 3:407-418 (1984); Rogers, J. Biol. Chem. 260:3731-3738 (1985); Rothstein et al., Gene 55:353-356 (1987); Whittier et al., Nucleic Acids Research 15:2515-2535 (1987); Wirsel et al., Molecular Microbiology 3:3-14 (1989); Yu et al., Gene 122:247-253 (1992). A description of the regulation of plant gene expression by the phytohormone, gibberellic acid and secreted enzymes induced by gibberellic acid can be found in R.L. Jones and J. MacMillin, Gibberellins: in: Advanced Plant Physiology, Malcolm B. Wilkins, ed., 1984 Pitman Publishing Limited, London, pp. 21-52.

References that describe other metabolically-regulated genes: Sheen, *Plant Cell*, 2:1027-1038(1990); Maas et al., *EMBO J.* 9:3447-3452 (1990); Benkel and Hickey, *Proc. Natl. Acad. Sci.* 84:1337-1339 (1987)

5

10

15

20

25

Typically, using techniques known in the art, a desired polynucleotide sequence is inserted into an expression cassette comprising genetic regulatory elements designed for operation in plants. The expression cassette is inserted into a desired expression vector with companion sequences upstream and downstream from the expression cassette suitable for expression in a plant host. The companion sequences will be of plasmid or viral origin and provide necessary characteristics to the vector to permit the vectors to move DNA from an original cloning host, such as bacteria, to the desired plant host. The basic bacterial/plant vector construct will preferably provide a broad host range prokaryote replication origin; a prokaryote selectable marker; and, for Agrobacterium transformations, T DNA sequences for Agrobacterium-mediated transfer to plant chromosomes. Where the heterologous gene is not readily amenable to detection, the construct will preferably also have a selectable marker gene suitable for determining if a plant cell has been transformed. A general review of suitable markers, for example for the members of the grass family, is found in Wilmink and Dons, 1993, *Plant Mol. Biol. Reptr.*, 11(2):165-185.

Sequences suitable for permitting integration of the heterologous sequence into the plant genome are also recommended. These might include transposon sequences and the like for homologous recombination as well as Ti sequences which permit random insertion of a heterologous expression cassette into a plant genome. Suitable prokaryote selectable markers include resistance toward antibiotics such as ampicillin or tetracycline. Other DNA sequences encoding additional functions may also be present in the vector, as is known in the art.

The nucleic acid molecules of the subject invention may be included into an expression cassette for expression of the protein(s) of interest. Usually, there will be only one expression cassette, although two or more are feasible. The recombinant expression cassette will contain in addition to the heterologous protein encoding sequence the following elements, a promoter region, plant 5' untranslated sequences, initiation codon depending upon whether or not the structural gene comes equipped with one, and a transcription and translation termination sequence. Unique restriction enzyme sites at the 5' and 3' ends of the cassette allow for easy insertion into a pre-existing vector.

A heterologous coding sequence may be for any protein relating to the present invention. The sequence encoding the protein of interest will encode a signal peptide which allows processing and translocation of the protein, as appropriate, and will usually lack any sequence which might result in the binding of the desired protein of the invention to a membrane. Since, for the most part, the transcriptional initiation region will be for a gene which is expressed and translocated during germination, by employing the signal peptide which provides for translocation, one may also provide for translocation of the protein of interest. In this way, the protein(s) of interest will be translocated from the cells in which they are expressed and may be efficiently harvested. Typically secretion in seeds are across the aleurone or scutellar epithelium layer into the endosperm of the seed. While it is not required that the protein be secreted from the cells in which the protein is produced, this facilitates the isolation and purification of the recombinant protein.

5

10

15

20

25

30

Since the ultimate expression of the desired gene product will be in a eucaryotic cell it is desirable to determine whether any portion of the cloned gene contains sequences which will be processed out as introns by the host's splicosome machinery. If so, site-directed mutagenesis of the "intron" region may be conducted to prevent losing a portion of the genetic message as a false intron code, Reed and Maniatis, *Cell* 41:95-105, 1985.

The vector can be microinjected directly into plant cells by use of micropipettes to mechanically transfer the recombinant DNA. Crossway, *Mol. Gen. Genet*, 202:179-185, 1985. The genetic material may also be transferred into the plant cell by using polyethylene glycol, Krens, et al., *Nature*, 296, 72-74, 1982. Another method of introduction of nucleic acid segments is high velocity ballistic penetration by small particles with the nucleic acid either within the matrix of small beads or particles, or on the surface, Klein, et al., *Nature*, 327, 70-73, 1987 and Knudsen and Muller, 1991, *Planta*, 185:330-336 teaching particle bombardment of barley endosperm to create transgenic barley. Yet another method of introduction would be fusion of protoplasts with other entities, either minicells, cells, lysosomes or other fusible lipid-surfaced bodies, Fraley, et al., *Proc. Natl. Acad. Sci. USA*, 79, 1859-1863, 1982.

The vector may also be introduced into the plant cells by electroporation. (Fromm et al., *Proc. Natl Acad. Sci. USA* 82:5824, 1985). In this technique, plant protoplasts are electroporated in the presence of plasmids containing the gene construct. Electrical impulses of high field strength reversibly permeabilize biomembranes allowing the introduction of the plasmids. Electroporated plant protoplasts reform the cell wall, divide, and form plant callus.

5

10

15

20

25

All plants from which protoplasts can be isolated and cultured to give whole regenerated plants can be transformed by the present invention so that whole plants are recovered which contain the transferred gene. It is known that practically all plants can be regenerated from cultured cells or tissues, including but not limited to all major species of sugarcane, sugar beet, cotton, fruit and other trees, legumes and vegetables. Some suitable plants include, for example, species from the genera Fragaria, Lotus, Medicago, Onobrychis, Trifolium, Trigonella, Vigna, Citrus, Linum, Geranium, Manihot, Daucus, Arabidopsis, Brassica, Raphanus, Sinapis, Atropa, Capsicum, Datura, Hyoscyamus, Lycopersion, Nicotiana, Solanum, Petunia, Digitalis, Majorana, Cichorium, Helianthus, Lactuca, Bromus, Asparagus, Antirrhinum, Hererocallis, Nemesia, Pelargonium, Panicum, Pennisetum, Ranunculus, Senecio, Salpiglossis, Cucumis, Browaalia, Glycine, Lolium, Zea, Triticum, Sorghum, and Datura.

Means for regeneration vary from species to species of plants, but generally a suspension of transformed protoplasts containing copies of the heterologous gene is first provided. Callus tissue is formed and shoots may be induced from callus and subsequently rooted. Alternatively, embryo formation can be induced from the protoplast suspension. These embryos germinate as natural embryos to form plants. The culture media will generally contain various amino acids and hormones, such as auxin and cytokinins. It is also advantageous to add glutamic acid and proline to the medium, especially for such species as corn and alfalfa. Shoots and roots normally develop simultaneously. Efficient regeneration will depend on the medium, on the genotype, and on the history of the culture. If these three variables are controlled, then regeneration is fully reproducible and repeatable.

In some plant cell culture systems, the desired protein of the invention may be excreted or alternatively, the protein may be extracted from the whole plant. Where the desired protein of the invention is secreted into the medium, it may be collected. Alternatively, the embryos and embryoless-half seeds or other plant tissue may be mechanically disrupted to release any secreted protein between cells and tissues. The mixture may be suspended in a buffer solution to retrieve soluble proteins. Conventional protein isolation and purification methods will be then used to purify the recombinant protein. Parameters of time, temperature pH, oxygen, and volumes will be adjusted through routine methods to optimize expression and recovery of heterologous protein.

# iv. Bacterial Systems

5

10

15

20

25

30

Bacterial expression techniques are known in the art. A bacterial promoter is any DNA sequence capable of binding bacterial RNA polymerase and initiating the downstream (3') transcription of a coding sequence (eg. structural gene) into mRNA. A promoter will have a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region usually includes an RNA polymerase binding site and a transcription initiation site. A bacterial promoter may also have a second domain called an operator, that may overlap an adjacent RNA polymerase binding site at which RNA synthesis begins. The operator permits negative regulated (inducible) transcription, as a gene repressor protein may bind the operator and thereby inhibit transcription of a specific gene. Constitutive expression may occur in the absence of negative regulatory elements, such as the operator. In addition, positive regulation may be achieved by a gene activator protein binding sequence, which, if present is usually proximal (5') to the RNA polymerase binding sequence. An example of a gene activator protein is the catabolite activator protein (CAP), which helps initiate transcription of the lac operon in Escherichia coli (E. coli) [Raibaud et al. (1984) Annu. Rev. Genet. 18:173]. Regulated expression may therefore be either positive or negative, thereby either enhancing or reducing transcription.

Sequences encoding metabolic pathway enzymes provide particularly useful promoter sequences. Examples include promoter sequences derived from sugar metabolizing enzymes, such as galactose, lactose (lac) [Chang et al. (1977) Nature 198:1056], and maltose. Additional examples include promoter sequences derived from biosynthetic enzymes such as tryptophan (trp) [Goeddel et al. (1980) Nuc. Acids Res. 8:4057; Yelverton et al. (1981) Nucl. Acids Res. 9:731; US patent 4,738,921; EP-A-0036776 and EP-A-0121775]. The g-laotamase (bla) promoter system [Weissmann (1981) "The cloning of interferon and other mistakes." In Interferon 3 (ed. I. Gresser)], bacteriophage lambda PL [Shimatake et al. (1981) Nature 292:128] and T5 [US patent 4,689,406] promoter systems also provide useful promoter sequences.

In addition, synthetic promoters which do not occur in nature also function as bacterial promoters. For example, transcription activation sequences of one bacterial or bacteriophage promoter may be joined with the operon sequences of another bacterial or bacteriophage promoter, creating a synthetic hybrid promoter [US patent 4,551,433]. For example, the *tac* promoter is a hybrid *trp-lac* promoter comprised of both *trp* promoter and *lac* operon sequences that is regulated by the *lac* repressor [Amann *et al.* (1983) *Gene 25*:167; de Boer *et al.* (1983) *Proc. Natl. Acad. Sci. 80*:21].

5

10

15

20

25

30

Furthermore, a bacterial promoter can include naturally occurring promoters of non-bacterial origin that have the ability to bind bacterial RNA polymerase and initiate transcription. A naturally occurring promoter of non-bacterial origin can also be coupled with a compatible RNA polymerase to produce high levels of expression of some genes in prokaryotes. The bacteriophage T7 RNA polymerase/promoter system is an example of a coupled promoter system [Studier et al. (1986) J. Mol. Biol. 189:113; Tabor et al. (1985) Proc Natl. Acad. Sci. 82:1074]. In addition, a hybrid promoter can also be comprised of a bacteriophage promoter and an E. coli operator region (EPO-A-0 267 851).

In addition to a functioning promoter sequence, an efficient ribosome binding site is also useful for the expression of foreign genes in prokaryotes. In *E. coli*, the ribosome binding site is called the Shine-Dalgarno (SD) sequence and includes an initiation codon (ATG) and a sequence 3-9 nucleotides in length located 3-11 nucleotides upstream of the initiation codon [Shine *et al.* (1975) *Nature 254*:34]. The SD sequence is thought to promote binding of mRNA to the ribosome by the pairing of bases between the SD sequence and the 3' and of *E. coli* 16S rRNA [Steitz *et al.* (1979) "Genetic signals and nucleotide sequences in messenger RNA." In *Biological Regulation and Development: Gene Expression* (ed. R.F. Goldberger)]. To express eukaryotic genes and prokaryotic genes with weak ribosome-binding site [Sambrook *et al.* (1989) "Expression of cloned genes in Escherichia coli." In *Molecular Cloning: A Laboratory Manual*].

A DNA molecule may be expressed intracellularly. A promoter sequence may be directly linked with the DNA molecule, in which case the first amino acid at the N-terminus will always be a methionine, which is encoded by the ATG start codon. If desired, methionine at the N-terminus may be cleaved from the protein by *in vitro* incubation with cyanogen bromide or by either *in vivo* on *in vitro* incubation with a bacterial methionine N-terminal peptidase (EPO-A-0 219 237).

Fusion proteins provide an alternative to direct expression. Usually, a DNA sequence encoding the N-terminal portion of an endogenous bacterial protein, or other stable protein, is fused to the 5' end of heterologous coding sequences. Upon expression, this construct will provide a fusion of the two amino acid sequences. For example, the bacteriophage lambda cell gene can be linked at the 5' terminus of a foreign gene and expressed in bacteria. The resulting fusion protein preferably retains a site for a processing enzyme (factor Xa) to cleave the bacteriophage protein from the foreign gene [Nagai et al. (1984) Nature 309:810]. Fusion proteins can also be made with sequences from the lacZ [Jia et al. (1987) Gene 60:197], trpE [Allen et al. (1987) J. Biotechnol. 5:93; Makoff et al.

(1989) J. Gen. Microbiol. 135:11], and Chey [EP-A-0 324 647] genes. The DNA sequence at the junction of the two amino acid sequences may or may not encode a cleavable site. Another example is a ubiquitin fusion protein. Such a fusion protein is made with the ubiquitin region that preferably retains a site for a processing enzyme (eg. ubiquitin specific processing-protease) to cleave the ubiquitin from the foreign protein. Through this method, native foreign protein can be isolated [Miller et al. (1989) Bio/Technology 7:698].

5

10

25

30

Alternatively, foreign proteins can also be secreted from the cell by creating chimeric DNA molecules that encode a fusion protein comprised of a signal peptide sequence fragment that provides for secretion of the foreign protein in bacteria [US patent 4,336,336]. The signal sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell. The protein is either secreted into the growth media (gram-positive bacteria) or into the periplasmic space, located between the inner and outer membrane of the cell (gram-negative bacteria). Preferably there are processing sites, which can be cleaved either *in vivo* or *in vitro* encoded between the signal peptide fragment and the foreign gene.

DNA encoding suitable signal sequences can be derived from genes for secreted bacterial proteins, such as the E. coli outer membrane protein gene (ompA) [Masui et al. (1983), in: Experimental Manipulation of Gene Expression; Ghrayeb et al. (1984) EMBO J. 3:2437] and the E. coli alkaline phosphatase signal sequence (phoA) [Oka et al. (1985) Proc. Natl. Acad. Sci. 82:7212]. As an additional example, the signal sequence of the alpha-amylase gene from various Bacillus strains can be used to secrete heterologous proteins from B. subtilis [Palva et al. (1982) Proc. Natl. Acad. Sci. USA 79:5582; EP-A-0 244 042].

Usually, transcription termination sequences recognized by bacteria are regulatory regions located 3' to the translation stop codon, and thus together with the promoter flank the coding sequence. These sequences direct the transcription of an mRNA which can be translated into the polypeptide encoded by the DNA. Transcription termination sequences frequently include DNA sequences of about 50 nucleotides capable of forming stem loop structures that aid in terminating transcription. Examples include transcription termination sequences derived from genes with strong promoters, such as the *trp* gene in *E. coli* as well as other biosynthetic genes.

Usually, the above described components, comprising a promoter, signal sequence (if desired), coding sequence of interest, and transcription termination sequence, are put together into expression constructs. Expression constructs are often maintained in a replicon, such as an extrachromosomal

5

10

15

20

element (eg. plasmids) capable of stable maintenance in a host, such as bacteria. The replicon will have a replication system, thus allowing it to be maintained in a prokaryotic host either for expression or for cloning and amplification. In addition, a replicon may be either a high or low copy number plasmid. A high copy number plasmid will generally have a copy number ranging from about 5 to about 200, and usually about 10 to about 150. A host containing a high copy number plasmid will preferably contain at least about 10, and more preferably at least about 20 plasmids. Either a high or low copy number vector may be selected, depending upon the effect of the vector and the foreign protein on the host.

Alternatively, the expression constructs can be integrated into the bacterial genome with an integrating vector. Integrating vectors usually contain at least one sequence homologous to the bacterial chromosome that allows the vector to integrate. Integrations appear to result from recombinations between homologous DNA in the vector and the bacterial chromosome. For example, integrating vectors constructed with DNA from various Bacillus strains integrate into the Bacillus chromosome (EP-A- 0 127 328). Integrating vectors may also be comprised of bacteriophage or transposon sequences.

Usually, extrachromosomal and integrating expression constructs may contain selectable markers to allow for the selection of bacterial strains that have been transformed. Selectable markers can be expressed in the bacterial host and may include genes which render bacteria resistant to drugs such as ampicillin, chloramphenicol, erythromycin, kanamycin (neomycin), and tetracycline [Davies et al. (1978) Annu. Rev. Microbiol. 32:469]. Selectable markers may also include biosynthetic genes, such as those in the histidine, tryptophan, and leucine biosynthetic pathways.

Alternatively, some of the above described components can be put together in transformation vectors. Transformation vectors are usually comprised of a selectable market that is either maintained in a replicon or developed into an integrating vector, as described above.

Expression and transformation vectors, either extra-chromosomal replicons or integrating vectors, have been developed for transformation into many bacteria. For example, expression vectors have been developed for, inter alia, the following bacteria: Bacillus subtilis [Palva et al. (1982) Proc. Natl. Acad. Sci. USA 79:5582; EP-A-0 036 259 and EP-A-0 063 953; WO 84/04541], Escherichia coli [Shimatake et al. (1981) Nature 292:128; Amann et al. (1985) Gene 40:183; Studier et al.
(1986) J. Mol. Biol. 189:113; EP-A-0 036 776,EP-A-0 136 829 and EP-A-0 136 907],

Streptococcus cremoris [Powell et al. (1988) Appl. Environ. Microbiol. 54:655]; Streptococcus lividans [Powell et al. (1988) Appl. Environ. Microbiol. 54:655], Streptomyces lividans [US patent 4,745,056].

Methods of introducing exogenous DNA into bacterial hosts are well-known in the art, and usually include either the transformation of bacteria treated with CaCl<sub>2</sub> or other agents, such as divalent cations and DMSO. DNA can also be introduced into bacterial cells by electroporation. Transformation procedures usually vary with the bacterial species to be transformed. See eg. [Masson et al. (1989) FEMS Microbiol. Lett. 60:273; Palva et al. (1982) Proc. Natl. Acad. Sci. USA 79:5582; EP-A-0 036 259 and EP-A-0 063 953; WO 84/04541, Bacillus], [Miller et al. (1988) Proc. Natl. Acad. Sci. 85:856; Wang et al. (1990) J. Bacteriol. 172:949, Campylobacter], [Cohen 10 et al. (1973) Proc. Natl. Acad. Sci. 69:2110; Dower et al. (1988) Nucleic Acids Res. 16:6127; Kushner (1978) "An improved method for transformation of Escherichia coli with ColE1-derived plasmids. In Genetic Engineering: Proceedings of the International Symposium on Genetic Engineering (eds. H.W. Boyer and S. Nicosia); Mandel et al. (1970) J. Mol. Biol. 53:159; Taketo (1988) Biochim. Biophys. Acta 949:318; Escherichia], [Chassy et al. (1987) FEMS Microbiol. Lett. 15 44:173 Lactobacillus]; [Fiedler et al. (1988) Anal. Biochem 170:38, Pseudomonas]; [Augustin et al. (1990) FEMS Microbiol. Lett. 66:203, Staphylococcus], [Barany et al. (1980) J. Bacteriol. 144:698; Harlander (1987) "Transformation of Streptococcus lactis by electroporation, in: Streptococcal Genetics (ed. J. Ferretti and R. Curtiss III); Perry et al. (1981) Infect. Immun. 32:1295; Powell et al. (1988) Appl. Environ. Microbiol. 54:655; Somkuti et al. (1987) Proc. 4th 20 Evr. Cong. Biotechnology 1:412, Streptococcus].

#### v. Yeast Expression

25

30

5

Yeast expression systems are also known to one of ordinary skill in the art. A yeast promoter is any DNA sequence capable of binding yeast RNA polymerase and initiating the downstream (3') transcription of a coding sequence (eg. structural gene) into mRNA. A promoter will have a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region usually includes an RNA polymerase binding site (the "TATA Box") and a transcription initiation site. A yeast promoter may also have a second domain called an upstream activator sequence (UAS), which, if present, is usually distal to the structural gene. The UAS permits regulated (inducible) expression. Constitutive expression occurs in the absence of a UAS. Regulated expression may be either positive or negative, thereby either enhancing or reducing transcription.

-23-

Yeast is a fermenting organism with an active metabolic pathway, therefore sequences encoding enzymes in the metabolic pathway provide particularly useful promoter sequences. Examples include alcohol dehydrogenase (ADH) (EP-A-0 284 044), enolase, glucokinase, glucose-6-phosphate isomerase, glyceraldehyde-3-phosphate-dehydrogenase (GAP or GAPDH), hexokinase, phosphofructokinase, 3-phosphoglycerate mutase, and pyruvate kinase (PyK) (EPO-A-0 329 203). The yeast *PHO5* gene, encoding acid phosphatase, also provides useful promoter sequences [Myanohara *et al.* (1983) *Proc. Natl. Acad. Sci. USA 80*:1].

5

- In addition, synthetic promoters which do not occur in nature also function as yeast promoters. For 10 example, UAS sequences of one yeast promoter may be joined with the transcription activation region of another yeast promoter, creating a synthetic hybrid promoter. Examples of such hybrid promoters include the ADH regulatory sequence linked to the GAP transcription activation region (US Patent Nos. 4,876,197 and 4,880,734). Other examples of hybrid promoters include promoters which consist of the regulatory sequences of either the ADH2, GAL4, GAL10, OR PHO5 genes, 15 combined with the transcriptional activation region of a glycolytic enzyme gene such as GAP or PyK (EP-A-0 164 556). Furthermore, a yeast promoter can include naturally occurring promoters of non-yeast origin that have the ability to bind yeast RNA polymerase and initiate transcription. Examples of such promoters include, inter alia, [Cohen et al. (1980) Proc. Natl. Acad. Sci. USA 77:1078; Henikoff et al. (1981) Nature 283:835; Hollenberg et al. (1981) Curr. Topics Microbiol. 20 Immunol. 96:119; Hollenberg et al. (1979) "The Expression of Bacterial Antibiotic Resistance Genes in the Yeast Saccharomyces cerevisiae," in: Plasmids of Medical, Environmental and Commercial Importance (eds. K.N. Timmis and A. Puhler); Mercerau-Puigalon et al. (1980) Gene 11:163; Panthier et al. (1980) Curr. Genet. 2:109;].
- A DNA molecule may be expressed intracellularly in yeast. A promoter sequence may be directly linked with the DNA molecule, in which case the first amino acid at the N-terminus of the recombinant protein will always be a methionine, which is encoded by the ATG start codon. If desired, methionine at the N-terminus may be cleaved from the protein by *in vitro* incubation with cyanogen bromide.

Fusion proteins provide an alternative for yeast expression systems, as well as in mammalian, baculovirus, and bacterial expression systems. Usually, a DNA sequence encoding the N-terminal portion of an endogenous yeast protein, or other stable protein, is fused to the 5' end of heterologous coding sequences. Upon expression, this construct will provide a fusion of the two amino acid sequences. For example, the yeast or human superoxide dismutase (SOD) gene, can be linked at the 5' terminus of a foreign gene and expressed in yeast. The DNA sequence at the junction of the two amino acid sequences may or may not encode a cleavable site. See eg. EP-A-0 196 056. Another example is a ubiquitin fusion protein. Such a fusion protein is made with the ubiquitin region that preferably retains a site for a processing enzyme (eg. ubiquitin-specific processing protease) to cleave the ubiquitin from the foreign protein. Through this method, therefore, native foreign protein can be isolated (eg. WO88/024066).

5

10

15

20

25

Alternatively, foreign proteins can also be secreted from the cell into the growth media by creating chimeric DNA molecules that encode a fusion protein comprised of a leader sequence fragment that provide for secretion in yeast of the foreign protein. Preferably, there are processing sites encoded between the leader fragment and the foreign gene that can be cleaved either *in vivo* or *in vitro*. The leader sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell.

DNA encoding suitable signal sequences can be derived from genes for secreted yeast proteins, such as the yeast invertase gene (EP-A-0 012 873; JPO. 62,096,086) and the A-factor gene (US patent 4,588,684). Alternatively, leaders of non-yeast origin, such as an interferon leader, exist that also provide for secretion in yeast (EP-A-0 060 057).

A preferred class of secretion leaders are those that employ a fragment of the yeast alpha-factor gene, which contains both a "pre" signal sequence, and a "pro" region. The types of alpha-factor fragments that can be employed include the full-length pre-pro alpha factor leader (about 83 amino acid residues) as well as truncated alpha-factor leaders (usually about 25 to about 50 amino acid residues) (US Patents 4,546,083 and 4,870,008; EP-A-0 324 274). Additional leaders employing an alpha-factor leader fragment that provides for secretion include hybrid alpha-factor leaders made with a presequence of a first yeast, but a pro-region from a second yeast alphafactor. (eg. see WO 89/02463.)

Usually, transcription termination sequences recognized by yeast are regulatory regions located 3' to the translation stop codon, and thus together with the promoter flank the coding sequence. These sequences direct the transcription of an mRNA which can be translated into the polypeptide encoded by the DNA. Examples of transcription terminator sequence and other yeast-recognized termination sequences, such as those coding for glycolytic enzymes.

5

10

15

Usually, the above described components, comprising a promoter, leader (if desired), coding sequence of interest, and transcription termination sequence, are put together into expression constructs. Expression constructs are often maintained in a replicon, such as an extrachromosomal element (eg. plasmids) capable of stable maintenance in a host, such as yeast or bacteria. The replicon may have two replication systems, thus allowing it to be maintained, for example, in yeast for expression and in a prokaryotic host for cloning and amplification. Examples of such yeast-bacteria shuttle vectors include YEp24 [Botstein et al. (1979) Gene 8:17-24], pCl/1 [Brake et al. (1984) Proc. Natl. Acad. Sci USA 81:4642-4646], and YRp17 [Stinchcomb et al. (1982) J. Mol. Biol. 158:157]. In addition, a replicon may be either a high or low copy number plasmid. A high copy number plasmid will generally have a copy number ranging from about 5 to about 200, and usually about 10 to about 150. A host containing a high copy number plasmid will preferably have at least about 10, and more preferably at least about 20. Enter a high or low copy number vector may be selected, depending upon the effect of the vector and the foreign protein on the host. See eg. Brake et al., supra.

Alternatively, the expression constructs can be integrated into the yeast genome with an integrating vector. Integrating vectors usually contain at least one sequence homologous to a yeast chromosome that allows the vector to integrate, and preferably contain two homologous sequences flanking the expression construct. Integrations appear to result from recombinations between homologous DNA in the vector and the yeast chromosome [Orr-Weaver et al. (1983) Methods in Enzymol. 101:228-245]. An integrating vector may be directed to a specific locus in yeast by selecting the appropriate homologous sequence for inclusion in the vector. See Orr-Weaver et al., supra. One or more expression construct may integrate, possibly affecting levels of recombinant protein produced [Rine et al. (1983) Proc. Natl. Acad. Sci. USA 80:6750]. The chromosomal sequences included in the vector can occur either as a single segment in the vector, which results in the integration of the entire vector, or two segments homologous to adjacent segments in the

chromosome and flanking the expression construct in the vector, which can result in the stable integration of only the expression construct.

Usually, extrachromosomal and integrating expression constructs may contain selectable markers to allow for the selection of yeast strains that have been transformed. Selectable markers may include biosynthetic genes that can be expressed in the yeast host, such as ADE2, HIS4, LEU2, TRP1, and ALG7, and the G418 resistance gene, which confer resistance in yeast cells to tunicamycin and G418, respectively. In addition, a suitable selectable marker may also provide yeast with the ability to grow in the presence of toxic compounds, such as metal. For example, the presence of CUPI allows yeast to grow in the presence of copper ions [Butt et al. (1987) Microbiol. Rev. 51:351].

5

10

15

20

30

Alternatively, some of the above described components can be put together into transformation vectors. Transformation vectors are usually comprised of a selectable marker that is either maintained in a replicon or developed into an integrating vector, as described above.

Expression and transformation vectors, either extrachromosomal replicons or integrating vectors, have been developed for transformation into many yeasts. For example, expression vectors have been developed for, inter alia, the following yeasts: Candida albicans [Kurtz, et al. (1986) Mol. Cell. Biol. 6:142], Candida maltosa [Kunze, et al. (1985) J. Basic Microbiol. 25:141]. Hansenula polymorpha [Gleeson, et al. (1986) J. Gen. Microbiol. 132:3459; Roggenkamp et al. (1986) Mol. Gen. Genet. 202:302], Kluyveromyces fragilis [Das, et al. (1984) J. Bacteriol. 158:1165], Kluyveromyces lactis [De Louvencourt et al. (1983) J. Bacteriol. 154:737; Van den Berg et al. (1990) Bio/Technology 8:135], Pichia guillerimondii [Kunze et al. (1985) J. Basic Microbiol. 25:141], Pichia pastoris [Cregg, et al. (1985) Mol. Cell. Biol. 5:3376; US Patent Nos. 4,837,148 and 4,929,555], Saccharomyces cerevisiae [Hinnen et al. (1978) Proc. Natl. Acad. Sci. USA 75:1929; Ito et al. (1983) J. Bacteriol. 153:163], Schizosaccharomyces pombe [Beach and Nurse 25 (1981) Nature 300:706], and Yarrowia lipolytica [Davidow, et al. (1985) Curr. Genet. 10:380471 Gaillardin, et al. (1985) Curr. Genet. 10:49].

Methods of introducing exogenous DNA into yeast hosts are well-known in the art, and usually include either the transformation of spheroplasts or of intact yeast cells treated with alkali cations. Transformation procedures usually vary with the yeast species to be transformed. See eg. [Kurtz et al. (1986) Mol. Cell. Biol. 6:142; Kunze et al. (1985) J. Basic Microbiol. 25:141; Candida]; [Gleeson et al. (1986) J. Gen. Microbiol. 132:3459; Roggenkamp et al. (1986) Mol. Gen. Genet. 202:302; Hansenula]; [Das et al. (1984) J. Bacteriol. 158:1165; De Louvencourt et al. (1983) J. Bacteriol. 154:1165; Van den Berg et al. (1990) Bio/Technology 8:135; Kluyveromyces]; [Cregg et al. (1985) Mol. Cell. Biol. 5:3376; Kunze et al. (1985) J. Basic Microbiol. 25:141; US Patent Nos. 4,837,148 and 4,929,555; Pichia]; [Hinnen et al. (1978) Proc. Natl. Acad. Sci. USA 75;1929; Ito et al. (1983) J. Bacteriol. 153:163 Saccharomyces]; [Beach and Nurse (1981) Nature 300:706; Schizosaccharomyces]; [Davidow et al. (1985) Curr. Genet. 10:39; Gaillardin et al. (1985) Curr. Genet. 10:49; Yarrowia].

#### **Antibodies**

5

20

25

30

As used herein, the term "antibody" refers to a polypeptide or group of polypeptides composed of at least one antibody combining site. An "antibody combining site" is the three-dimensional binding space with an internal surface shape and charge distribution complementary to the features of an epitope of an antigen, which allows a binding of the antibody with the antigen. "Antibody" includes, for example, vertebrate antibodies, hybrid antibodies, chimeric antibodies, humanised antibodies, altered antibodies, univalent antibodies, Fab proteins, and single domain antibodies.

Antibodies against the proteins of the invention are useful for affinity chromatography, immunoassays, and distinguishing/identifying Neisserial proteins.

Antibodies to the proteins of the invention, both polyclonal and monoclonal, may be prepared by conventional methods. In general, the protein is first used to immunize a suitable animal, preferably a mouse, rat, rabbit or goat. Rabbits and goats are preferred for the preparation of polyclonal sera due to the volume of serum obtainable, and the availability of labeled anti-rabbit and anti-goat antibodies. Immunization is generally performed by mixing or emulsifying the protein in saline, preferably in an adjuvant such as Freund's complete adjuvant, and injecting the mixture or emulsion parenterally (generally subcutaneously or intramuscularly). A dose of 50-200 µg/injection is typically sufficient. Immunization is generally boosted 2-6 weeks later with one or more injections of the protein in saline, preferably using Freund's incomplete adjuvant. One may alternatively generate antibodies by in vitro immunization using methods known in the art, which for the purposes of this invention is considered equivalent to *in vivo* immunization. Polyclonal antisera is obtained by bleeding the immunized animal into a glass or plastic container, incubating the blood at 25°C for one hour, followed by incubating at 4°C for 2-18 hours. The serum is

recovered by centrifugation (eg. 1,000g for 10 minutes). About 20-50 ml per bleed may be obtained from rabbits.

5

10

15

20

25

30

Monoclonal antibodies are prepared using the standard method of Kohler & Milstein [Nature (1975) 256:495-96], or a modification thereof. Typically, a mouse or rat is immunized as described above. However, rather than bleeding the animal to extract serum, the spleen (and optionally several large lymph nodes) is removed and dissociated into single cells. If desired, the spleen cells may be screened (after removal of nonspecifically adherent cells) by applying a cell suspension to a plate or well coated with the protein antigen. B-cells expressing membrane-bound immunoglobulin specific for the antigen bind to the plate, and are not rinsed away with the rest of the suspension. Resulting B-cells, or all dissociated spleen cells, are then induced to fuse with myeloma cells to form hybridomas, and are cultured in a selective medium (eg. hypoxanthine, aminopterin, thymidine medium, "HAT"). The resulting hybridomas are plated by limiting dilution, and are assayed for the production of antibodies which bind specifically to the immunizing antigen (and which do not bind to unrelated antigens). The selected MAb-secreting hybridomas are then cultured either in vitro (eg. in tissue culture bottles or hollow fiber reactors), or in vivo (as ascites in mice).

If desired, the antibodies (whether polyclonal or monoclonal) may be labeled using conventional techniques. Suitable labels include fluorophores, chromophores, radioactive atoms (particularly <sup>32</sup>P and <sup>125</sup>I), electron-dense reagents, enzymes, and ligands having specific binding partners. Enzymes are typically detected by their activity. For example, horseradish peroxidase is usually detected by its ability to convert 3,3',5,5'-tetramethylbenzidine (TMB) to a blue pigment, quantifiable with a spectrophotometer. "Specific binding partner" refers to a protein capable of binding a ligand molecule with high specificity, as for example in the case of an antigen and a monoclonal antibody specific therefor. Other specific binding partners include biotin and avidin or streptavidin, IgG and protein A, and the numerous receptor-ligand couples known in the art. It should be understood that the above description is not meant to categorize the various labels into distinct classes, as the same label may serve in several different modes. For example, <sup>125</sup>I may serve as a radioactive label or as an electron-dense reagent. HRP may serve as enzyme or as antigen for a MAb. Further, one may combine various labels for desired effect. For example, MAbs and avidin also require labels in the practice of this invention: thus, one might label a MAb with biotin, and detect its presence with avidin labeled with <sup>125</sup>I, or with an anti-biotin MAb labeled with HRP. Other permutations and possibilities will be

readily apparent to those of ordinary skill in the art, and are considered as equivalents within the scope of the instant invention.

#### Pharmaceutical Compositions

5

10

15

20

25

30

Pharmaceutical compositions can comprise either polypeptides, antibodies, or nucleic acid of the invention. The pharmaceutical compositions will comprise a therapeutically effective amount of either polypeptides, antibodies, or polynucleotides of the claimed invention.

The term "therapeutically effective amount" as used herein refers to an amount of a therapeutic agent to treat, ameliorate, or prevent a desired disease or condition, or to exhibit a detectable therapeutic or preventative effect. The effect can be detected by, for example, chemical markers or antigen levels. Therapeutic effects also include reduction in physical symptoms, such as decreased body temperature. The precise effective amount for a subject will depend upon the subject's size and health, the nature and extent of the condition, and the therapeutics or combination of therapeutics selected for administration. Thus, it is not useful to specify an exact effective amount in advance. However, the effective amount for a given situation can be determined by routine experimentation and is within the judgement of the clinician.

For purposes of the present invention, an effective dose will be from about 0.01 mg/kg to 50 mg/kg or 0.05 mg/kg to about 10 mg/kg of the DNA constructs in the individual to which it is administered.

A pharmaceutical composition can also contain a pharmaceutically acceptable carrier. The term "pharmaceutically acceptable carrier" refers to a carrier for administration of a therapeutic agent, such as antibodies or a polypeptide, genes, and other therapeutic agents. The term refers to any pharmaceutical carrier that does not itself induce the production of antibodies harmful to the individual receiving the composition, and which may be administered without undue toxicity. Suitable carriers may be large, slowly metabolized macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers, and inactive virus particles. Such carriers are well known to those of ordinary skill in the art.

Pharmaceutically acceptable salts can be used therein, for example, mineral acid salts such as hydrochlorides, hydrobromides, phosphates, sulfates, and the like; and the salts of organic acids such as acetates, propionates, malonates, benzoates, and the like. A thorough discussion of pharmaceutically acceptable excipients is available in Remington's Pharmaceutical Sciences (Mack Pub. Co., N.J. 1991).

Pharmaceutically acceptable carriers in therapeutic compositions may contain liquids such as water, saline, glycerol and ethanol. Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, and the like, may be present in such vehicles. Typically, the therapeutic compositions are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid vehicles prior to injection may also be prepared. Liposomes are included within the definition of a pharmaceutically acceptable carrier.

### Delivery Methods

5

Once formulated, the compositions of the invention can be administered directly to the subject. The subjects to be treated can be animals; in particular, human subjects can be treated.

Direct delivery of the compositions will generally be accomplished by injection, either subcutaneously, intraperitoneally, intravenously or intramuscularly or delivered to the interstitial space of a tissue. The compositions can also be administered into a lesion. Other modes of administration include oral and pulmonary administration, suppositories, and transdermal or transcutaneous applications (eg. see WO98/20734), needles, and gene guns or hyposprays. Dosage treatment may be a single dose schedule or a multiple dose schedule.

#### Vaccines

20

25

30

Vaccines according to the invention may either be prophylactic (ie. to prevent infection) or therapeutic (ie. to treat disease after infection).

Such vaccines comprise immunising antigen(s), immunogen(s), polypeptide(s), protein(s) or nucleic acid, usually in combination with "pharmaceutically acceptable carriers," which include any carrier that does not itself induce the production of antibodies harmful to the individual receiving the composition. Suitable carriers are typically large, slowly metabolized macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers, lipid aggregates (such as oil droplets or liposomes), and inactive virus particles. Such carriers are well known to those of ordinary skill in the art. Additionally, these carriers may function as immunostimulating agents ("adjuvants"). Furthermore, the antigen or immunogen may be conjugated to a bacterial toxoid, such as a toxoid from diphtheria, tetanus, cholera, *H. pylori*, *etc.* pathogens.

Preferred adjuvants to enhance effectiveness of the composition include, but are not limited to: (1) aluminum salts (alum), such as aluminum hydroxide, aluminum phosphate, aluminum sulfate, etc; (2) oil-in-water emulsion formulations (with or without other specific immunostimulating agents

such as muramyl peptides (see below) or bacterial cell wall components), such as for example (a) MF59<sup>TM</sup> (WO 90/14837; Chapter 10 in Vaccine design: the subunit and adjuvant approach, eds. Powell & Newman, Plenum Press 1995), containing 5% Squalene, 0.5% Tween 80, and 0.5% Span 85 (optionally containing various amounts of MTP-PE (see below), although not required) formulated into submicron particles using a microfluidizer such as Model 110Y microfluidizer (Microfluidics, Newton, MA), (b) SAF, containing 10% Squalane, 0.4% Tween 80, 5% pluronicblocked polymer L121, and thr-MDP (see below) either microfluidized into a submicron emulsion or vortexed to generate a larger particle size emulsion, and (c) Ribi™ adjuvant system (RAS), (Ribi Immunochem, Hamilton, MT) containing 2% Squalene, 0.2% Tween 80, and one or more bacterial cell wall components from the group consisting of monophosphorylipid A (MPL), trehalose dimycolate (TDM), and cell wall skeleton (CWS), preferably MPL + CWS (Detox™); (3) saponin adjuvants, such as Stimulon™ (Cambridge Bioscience, Worcester, MA) may be used or particles generated therefrom such as ISCOMs (immunostimulating complexes); (4) Complete Freund's Adjuvant (CFA) and Incomplete Freund's Adjuvant (IFA); (5) cytokines, such as interleukins (eg. IL-1, IL-2, IL-4, IL-5, IL-6, IL-7, IL-12, etc.), interferons (eg. gamma interferon), macrophage colony stimulating factor (M-CSF), tumor necrosis factor (TNF), etc; and (6) other substances that act as immunostimulating agents to enhance the effectiveness of the composition. Alum and MF59<sup>TM</sup> are preferred.

5

10

15

25

30

As mentioned above, muramyl peptides include, but are not limited to, N-acetyl-muramyl-L-20 threonyl-D-isoglutamine (thr-MDP), N-acetyl-normuramyl-L-alanyl-D-isoglutamine (nor-MDP), N-acetylmuramyl-L-alanyl-D-isoglutaminyl-L-alanine-2-(1'-2'-dipalmitoyl-sn-glycero-3-hydroxyphosphoryloxy)-ethylamine (MTP-PE), etc.

The immunogenic compositions (eg. the immunising antigen/immunogen/polypeptide/protein/nucleic acid, pharmaceutically acceptable carrier, and adjuvant) typically will contain diluents, such as water, saline, glycerol, ethanol, etc. Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, and the like, may be present in such vehicles.

Typically, the immunogenic compositions are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid vehicles prior to injection may also be prepared. The preparation also may be emulsified or encapsulated in liposomes for enhanced adjuvant effect, as discussed above under pharmaceutically acceptable carriers.

Immunogenic compositions used as vaccines comprise an immunologically effective amount of the antigenic or immunogenic polypeptides, as well as any other of the above-mentioned components, as needed. By "immunologically effective amount", it is meant that the administration of that amount to an individual, either in a single dose or as part of a series, is effective for treatment or prevention. This amount varies depending upon the health and physical condition of the individual to be treated, the taxonomic group of individual to be treated (eg. nonhuman primate, primate, etc.), the capacity of the individual's immune system to synthesize antibodies, the degree of protection desired, the formulation of the vaccine, the treating doctor's assessment of the medical situation, and other relevant factors. It is expected that the amount will fall in a relatively broad range that can be determined through routine trials.

The immunogenic compositions are conventionally administered parenterally, eg. by injection, either subcutaneously, intramuscularly, or transdermally/transcutaneously (eg. WO98/20734). Additional formulations suitable for other modes of administration include oral and pulmonary formulations, suppositories, and transdermal applications. Dosage treatment may be a single dose schedule or a multiple dose schedule. The vaccine may be administered in conjunction with other immunoregulatory agents.

As an alternative to protein-based vaccines, DNA vaccination may be employed [eg. Robinson & Torres (1997) Seminars in Immunology 9:271-283; Donnelly et al. (1997) Annu Rev Immunol 15:617-648; see later herein].

#### 20 Gene Delivery Vehicles

5

10

15

25

30

Gene therapy vehicles for delivery of constructs including a coding sequence of a therapeutic of the invention, to be delivered to the mammal for expression in the mammal, can be administered either locally or systemically. These constructs can utilize viral or non-viral vector approaches in *in vivo* or *ex vivo* modality. Expression of such coding sequence can be induced using endogenous mammalian or heterologous promoters. Expression of the coding sequence in vivo can be either constitutive or regulated.

The invention includes gene delivery vehicles capable of expressing the contemplated nucleic acid sequences. The gene delivery vehicle is preferably a viral vector and, more preferably, a retroviral, adenoviral, adeno-associated viral (AAV), herpes viral, or alphavirus vector. The viral vector can also be an astrovirus, coronavirus, orthomyxovirus, papovavirus, paramyxovirus, parvovirus,

picornavirus, poxvirus, or togavirus viral vector. See generally, Jolly (1994) Cancer Gene Therapy 1:51-64; Kimura (1994) Human Gene Therapy 5:845-852; Connelly (1995) Human Gene Therapy 6:185-193; and Kaplitt (1994) Nature Genetics 6:148-153.

Retroviral vectors are well known in the art and we contemplate that any retroviral gene therapy vector is employable in the invention, including B, C and D type retroviruses, xenotropic retroviruses (for example, NZB-X1, NZB-X2 and NZB9-1 (see O'Neill (1985) *J. Virol.* 53:160) polytropic retroviruses eg. MCF and MCF-MLV (see Kelly (1983) *J. Virol.* 45:291), spumaviruses and lentiviruses. See RNA Tumor Viruses, Second Edition, Cold Spring Harbor Laboratory, 1985.

5

10

15

20

25

Portions of the retroviral gene therapy vector may be derived from different retroviruses. For example, retrovector LTRs may be derived from a Murine Sarcoma Virus, a tRNA binding site from a Rous Sarcoma Virus, a packaging signal from a Murine Leukemia Virus, and an origin of second strand synthesis from an Avian Leukosis Virus.

These recombinant retroviral vectors may be used to generate transduction competent retroviral vector particles by introducing them into appropriate packaging cell lines (see US patent 5,591,624). Retrovirus vectors can be constructed for site-specific integration into host cell DNA by incorporation of a chimeric integrase enzyme into the retroviral particle (see WO96/37626). It is preferable that the recombinant viral vector is a replication defective recombinant virus.

Packaging cell lines suitable for use with the above-described retrovirus vectors are well known in the art, are readily prepared (see WO95/30763 and WO92/05266), and can be used to create producer cell lines (also termed vector cell lines or "VCLs") for the production of recombinant vector particles. Preferably, the packaging cell lines are made from human parent cells (eg. HT1080 cells) or mink parent cell lines, which eliminates inactivation in human serum.

Preferred retroviruses for the construction of retroviral gene therapy vectors include Avian Leukosis Virus, Bovine Leukemia, Virus, Murine Leukemia Virus, Mink-Cell Focus-Inducing Virus, Murine Sarcoma Virus, Reticuloendotheliosis Virus and Rous Sarcoma Virus. Particularly preferred Murine Leukemia Viruses include 4070A and 1504A (Hartley and Rowe (1976) *J Virol* 19:19-25), Abelson (ATCC No. VR-999), Friend (ATCC No. VR-245), Graffi, Gross (ATCC Nol VR-590), Kirsten, Harvey Sarcoma Virus and Rauscher (ATCC No. VR-998) and Moloney Murine Leukemia Virus (ATCC No. VR-190). Such retroviruses may be obtained from depositories or

collections such as the American Type Culture Collection ("ATCC") in Rockville, Maryland or isolated from known sources using commonly available techniques.

Exemplary known retroviral gene therapy vectors employable in this invention include those described in patent applications GB2200651, EP0415731, EP0345242, EP0334301, WO89/02468; WO89/05349, WO89/09271, WO90/02806, WO90/07936, WO94/03622, WO93/25698, WO93/25234, WO93/11230, WO93/10218, WO91/02805, WO91/02825, WO95/07994, US 5,219,740, US 4,405,712, US 4,861,719, US 4,980,289, US 4,777,127, US 5,591,624. See also Vile (1993) Cancer Res 53:3860-3864; Vile (1993) Cancer Res 53:962-967; Ram (1993) Cancer Res 53 (1993) 83-88; Takamiya (1992) J Neurosci Res 33:493-503; Baba (1993) J Neurosurg 79:729-735; Mann (1983) Cell 33:153; Cane (1984) Proc Natl Acad Sci 81:6349; and Miller (1990) Human Gene Therapy 1.

5

10

15

20

25

30

Human adenoviral gene therapy vectors are also known in the art and employable in this invention. See, for example, Berkner (1988) Biotechniques 6:616 and Rosenfeld (1991) Science 252:431, and WO93/07283, WO93/06223, and WO93/07282. Exemplary known adenoviral gene therapy vectors employable in this invention include those described in the above referenced documents and in WO94/12649, WO93/03769, WO93/19191, WO94/28938, WO95/11984, WO95/00655, WO95/27071, WO95/29993, WO95/34671, WO96/05320, WO94/08026, WO94/11506, WO93/06223, WO94/24299, WO95/14102, WO95/24297, WO95/02697, WO94/28152, WO94/24299, WO95/09241, WO95/25807, WO95/05835, WO94/18922 and WO95/09654. Alternatively, administration of DNA linked to killed adenovirus as described in Curiel (1992) Hum. Gene Ther. 3:147-154 may be employed. The gene delivery vehicles of the invention also include adenovirus associated virus (AAV) vectors. Leading and preferred examples of such vectors for use in this invention are the AAV-2 based vectors disclosed in Srivastava, WO93/09239. Most preferred AAV vectors comprise the two AAV inverted terminal repeats in which the native D-sequences are modified by substitution of nucleotides, such that at least 5 native nucleotides and up to 18 native nucleotides, preferably at least 10 native nucleotides up to 18 native nucleotides, most preferably 10 native nucleotides are retained and the remaining nucleotides of the D-sequence are deleted or replaced with non-native nucleotides. The native D-sequences of the AAV inverted terminal repeats are sequences of 20 consecutive nucleotides in each AAV inverted terminal repeat (ie. there is one sequence at each end) which are not involved in HP formation. The non-native replacement nucleotide may be any nucleotide other than the nucleotide found in the

native D-sequence in the same position. Other employable exemplary AAV vectors are pWP-19, pWN-1, both of which are disclosed in Nahreini (1993) *Gene* 124:257-262. Another example of such an AAV vector is psub201 (see Samulski (1987) *J. Virol.* 61:3096). Another exemplary AAV vector is the Double-D ITR vector. Construction of the Double-D ITR vector is disclosed in US Patent 5,478,745. Still other vectors are those disclosed in Carter US Patent 4,797,368 and Muzyczka US Patent 5,139,941, Chartejee US Patent 5,474,935, and Kotin WO94/288157. Yet a further example of an AAV vector employable in this invention is SSV9AFABTKneo, which contains the AFP enhancer and albumin promoter and directs expression predominantly in the liver. Its structure and construction are disclosed in Su (1996) *Human Gene Therapy* 7:463-470. Additional AAV gene therapy vectors are described in US 5,354,678, US 5,173,414, US 5,139,941, and US 5,252,479.

The gene therapy vectors of the invention also include herpes vectors. Leading and preferred examples are herpes simplex virus vectors containing a sequence encoding a thymidine kinase polypeptide such as those disclosed in US 5,288,641 and EP0176170 (Roizman). Additional exemplary herpes simplex virus vectors include HFEM/ICP6-LacZ disclosed in WO95/04139 (Wistar Institute), pHSVlac described in Geller (1988) *Science* 241:1667-1669 and in WO90/09441 and WO92/07945, HSV Us3::pgC-lacZ described in Fink (1992) *Human Gene Therapy* 3:11-19 and HSV 7134, 2 RH 105 and GAL4 described in EP 0453242 (Breakefield), and those deposited with the ATCC as accession numbers ATCC VR-977 and ATCC VR-260.

15

Also contemplated are alpha virus gene therapy vectors that can be employed in this invention. Preferred alpha virus vectors are Sindbis viruses vectors. Togaviruses, Semliki Forest virus (ATCC VR-67; ATCC VR-1247), Middleberg virus (ATCC VR-370), Ross River virus (ATCC VR-373; ATCC VR-1246), Venezuelan equine encephalitis virus (ATCC VR923; ATCC VR-1250; ATCC VR-1249; ATCC VR-532), and those described in US patents 5,091,309, 5,217,879, and WO92/10578. More particularly, those alpha virus vectors described in US Serial No. 08/405,627, filed March 15, 1995,WO94/21792, WO92/10578, WO95/07994, US 5,091,309 and US 5,217,879 are employable. Such alpha viruses may be obtained from depositories or collections such as the ATCC in Rockville, Maryland or isolated from known sources using commonly available techniques. Preferably, alphavirus vectors with reduced cytotoxicity are used (see USSN 08/679640).

DNA vector systems such as eukarytic layered expression systems are also useful for expressing the nucleic acids of the invention. See WO95/07994 for a detailed description of eukaryotic layered expression systems. Preferably, the eukaryotic layered expression systems of the invention are derived from alphavirus vectors and most preferably from Sindbis viral vectors.

PCT/IB98/01665

Other viral vectors suitable for use in the present invention include those derived from poliovirus, for 5 example ATCC VR-58 and those described in Evans, Nature 339 (1989) 385 and Sabin (1973) J. Biol. Standardization 1:115; rhinovirus, for example ATCC VR-1110 and those described in Arnold (1990) J Cell Biochem L401; pox viruses such as canary pox virus or vaccinia virus, for example ATCC VR-111 and ATCC VR-2010 and those described in Fisher-Hoch (1989) Proc Natl Acad Sci 86:317: Flexner (1989) Ann NY Acad Sci 569:86, Flexner (1990) Vaccine 8:17; in US 4,603,112 and US 10 4.769.330 and WO89/01973; SV40 virus, for example ATCC VR-305 and those described in Mulligan (1979) Nature 277:108 and Madzak (1992) J Gen Virol 73:1533; influenza virus, for example ATCC VR-797 and recombinant influenza viruses made employing reverse genetics techniques as described in US 5,166,057 and in Enami (1990) Proc Natl Acad Sci 87:3802-3805; Enami & Palese (1991) J Virol 65:2711-2713 and Luytjes (1989) Cell 59:110, (see also McMichael 15 (1983) NEJ Med 309:13, and Yap (1978) Nature 273:238 and Nature (1979) 277:108); human immunodeficiency virus as described in EP-0386882 and in Buchschacher (1992) J. Virol. 66:2731; measles virus, for example ATCC VR-67 and VR-1247 and those described in EP-0440219; Aura virus, for example ATCC VR-368; Bebaru virus, for example ATCC VR-600 and ATCC VR-1240; Cabassou virus, for example ATCC VR-922; Chikungunya virus, for example ATCC VR-64 and 20 ATCC VR-1241; Fort Morgan Virus, for example ATCC VR-924; Getah virus, for example ATCC VR-369 and ATCC VR-1243; Kyzylagach virus, for example ATCC VR-927; Mayaro virus, for example ATCC VR-66; Mucambo virus, for example ATCC VR-580 and ATCC VR-1244; Ndumu virus, for example ATCC VR-371; Pixuna virus, for example ATCC VR-372 and ATCC VR-1245; Tonate virus, for example ATCC VR-925; Triniti virus, for example ATCC VR-469; Una virus, for 25 example ATCC VR-374; Whataroa virus, for example ATCC VR-926; Y-62-33 virus, for example ATCC VR-375; O'Nyong virus, Eastern encephalitis virus, for example ATCC VR-65 and ATCC VR-1242; Western encephalitis virus, for example ATCC VR-70, ATCC VR-1251, ATCC VR-622 and ATCC VR-1252; and coronavirus, for example ATCC VR-740 and those described in Hamre (1966) Proc Soc Exp Biol Med 121:190. 30

Delivery of the compositions of this invention into cells is not limited to the above mentioned viral vectors. Other delivery methods and media may be employed such as, for example, nucleic acid

expression vectors, polycationic condensed DNA linked or unlinked to killed adenovirus alone, for example see US Serial No. 08/366,787, filed December 30, 1994 and Curiel (1992) *Hum Gene Ther* 3:147-154 ligand linked DNA, for example see Wu (1989) *J Biol Chem* 264:16985-16987, eucaryotic cell delivery vehicles cells, for example see US Serial No.08/240,030, filed May 9, 1994, and US Serial No. 08/404,796, deposition of photopolymerized hydrogel materials, hand-held gene transfer particle gun, as described in US Patent 5,149,655, ionizing radiation as described in US5,206,152 and in WO92/11033, nucleic charge neutralization or fusion with cell membranes. Additional approaches are described in Philip (1994) *Mol Cell Biol* 14:2411-2418 and in Woffendin (1994) *Proc Natl Acad Sci* 91:1581-1585.

5

20

25

30

Particle mediated gene transfer may be employed, for example see US Serial No. 60/023,867. Briefly, the sequence can be inserted into conventional vectors that contain conventional control sequences for high level expression, and then incubated with synthetic gene transfer molecules such as polymeric DNA-binding cations like polylysine, protamine, and albumin, linked to cell targeting ligands such as asialoorosomucoid, as described in Wu & Wu (1987) J. Biol. Chem.

262:4429-4432, insulin as described in Hucked (1990) Biochem Pharmacol 40:253-263, galactose as described in Plank (1992) Bioconjugate Chem 3:533-539, lactose or transferrin.

Naked DNA may also be employed. Exemplary naked DNA introduction methods are described in WO 90/11092 and US 5,580,859. Uptake efficiency may be improved using biodegradable latex beads. DNA coated latex beads are efficiently transported into cells after endocytosis initiation by the beads. The method may be improved further by treatment of the beads to increase hydrophobicity and thereby facilitate disruption of the endosome and release of the DNA into the cytoplasm.

Liposomes that can act as gene delivery vehicles are described in US 5,422,120, WO95/13796, WO94/23697, WO91/14445 and EP-524,968. As described in USSN. 60/023,867, on non-viral delivery, the nucleic acid sequences encoding a polypeptide can be inserted into conventional vectors that contain conventional control sequences for high level expression, and then be incubated with synthetic gene transfer molecules such as polymeric DNA-binding cations like polylysine, protamine, and albumin, linked to cell targeting ligands such as asialoorosomucoid, insulin, galactose, lactose, or transferrin. Other delivery systems include the use of liposomes to encapsulate DNA comprising the gene under the control of a variety of tissue-specific or ubiquitously-active promoters. Further non-viral delivery suitable for use includes mechanical delivery systems such as the approach described in Woffendin et al (1994) Proc. Natl. Acad. Sci. USA

91(24):11581-11585. Moreover, the coding sequence and the product of expression of such can be delivered through deposition of photopolymerized hydrogel materials. Other conventional methods for gene delivery that can be used for delivery of the coding sequence include, for example, use of hand-held gene transfer particle gun, as described in US 5,149,655; use of ionizing radiation for activating transferred gene, as described in US 5,206,152 and WO92/11033

Exemplary liposome and polycationic gene delivery vehicles are those described in US 5,422,120 and 4,762,915; inWO 95/13796; WO94/23697; and WO91/14445; in EP-0524968; and in Stryer, Biochemistry, pages 236-240 (1975) W.H. Freeman, San Francisco; Szoka (1980) Biochem Biophys Acta 600:1; Bayer (1979) Biochem Biophys Acta 550:464; Rivnay (1987) Meth Enzymol 149:119; Wang (1987) Proc Natl Acad Sci 84:7851; Plant (1989) Anal Biochem 176:420.

A polynucleotide composition can comprises therapeutically effective amount of a gene therapy vehicle, as the term is defined above. For purposes of the present invention, an effective dose will be from about 0.01 mg/kg to 50 mg/kg or 0.05 mg/kg to about 10 mg/kg of the DNA constructs in the individual to which it is administered.

#### 15 <u>Delivery Methods</u>

5

10

Once formulated, the polynucleotide compositions of the invention can be administered (1) directly to the subject; (2) delivered ex vivo, to cells derived from the subject; or (3) in vitro for expression of recombinant proteins. The subjects to be treated can be mammals or birds. Also, human subjects can be treated.

Direct delivery of the compositions will generally be accomplished by injection, either subcutaneously, intraperitoneally, intravenously or intramuscularly or delivered to the interstitial space of a tissue. The compositions can also be administered into a lesion. Other modes of administration include oral and pulmonary administration, suppositories, and transdermal or transcutaneous applications (eg. see WO98/20734), needles, and gene guns or hyposprays. Dosage treatment may be a single dose schedule or a multiple dose schedule.

Methods for the ex vivo delivery and reimplantation of transformed cells into a subject are known in the art and described in eg. WO93/14778. Examples of cells useful in ex vivo applications include, for example, stem cells, particularly hematopoetic, lymph cells, macrophages, dendritic cells, or tumor cells.

Generally, delivery of nucleic acids for both ex vivo and in vitro applications can be accomplished by the following procedures, for example, dextran-mediated transfection, calcium phosphate precipitation, polybrene mediated transfection, protoplast fusion, electroporation, encapsulation of the polynucleotide(s) in liposomes, and direct microinjection of the DNA into nuclei, all well known in the art.

# Polynucleotide and polypeptide pharmaceutical compositions

In addition to the pharmaceutically acceptable carriers and salts described above, the following additional agents can be used with polynucleotide and/or polypeptide compositions.

#### A.Polypeptides

5

One example are polypeptides which include, without limitation: asioloorosomucoid (ASOR); transferrin; asialoglycoproteins; antibodies; antibody fragments; ferritin; interleukins; interferons, granulocyte, macrophage colony stimulating factor (GM-CSF), granulocyte colony stimulating factor (G-CSF), macrophage colony stimulating factor (M-CSF), stem cell factor and erythropoietin. Viral antigens, such as envelope proteins, can also be used. Also, proteins from other invasive organisms, such as the 17 amino acid peptide from the circumsporozoite protein of plasmodium falciparum known as RII.

## B.Hormones, Vitamins, etc.

Other groups that can be included are, for example: hormones, steroids, androgens, estrogens, thyroid hormone, or vitamins, folic acid.

#### 20 C.Polyalkylenes, Polysaccharides, etc.

Also, polyalkylene glycol can be included with the desired polynucleotides/polypeptides. In a preferred embodiment, the polyalkylene glycol is polyethlylene glycol. In addition, mono-, di-, or polysaccarides can be included. In a preferred embodiment of this aspect, the polysaccharide is dextran or DEAE-dextran. Also, chitosan and poly(lactide-co-glycolide)

#### 25 D.Lipids, and Liposomes

30

The desired polynucleotide/polypeptide can also be encapsulated in lipids or packaged in liposomes prior to delivery to the subject or to cells derived therefrom.

Lipid encapsulation is generally accomplished using liposomes which are able to stably bind or entrap and retain nucleic acid. The ratio of condensed polynucleotide to lipid preparation can vary but will generally be around 1:1 (mg DNA:micromoles lipid), or more of lipid. For a review of the

10

25

use of liposomes as carriers for delivery of nucleic acids, see, Hug and Sleight (1991) *Biochim. Biophys. Acta.* 1097:1-17; Straubinger (1983) *Meth. Enzymol.* 101:512-527.

Liposomal preparations for use in the present invention include cationic (positively charged), anionic (negatively charged) and neutral preparations. Cationic liposomes have been shown to mediate intracellular delivery of plasmid DNA (Felgner (1987) *Proc. Natl. Acad. Sci. USA* 84:7413-7416); mRNA (Malone (1989) *Proc. Natl. Acad. Sci. USA* 86:6077-6081); and purified transcription factors (Debs (1990) *J. Biol. Chem.* 265:10189-10192), in functional form.

Cationic liposomes are readily available. For example, N[1-2,3-dioleyloxy)propyl]-N,N,N-triethylammonium (DOTMA) liposomes are available under the trademark Lipofectin, from GIBCO BRL, Grand Island, NY. (See, also, Felgner *supra*). Other commercially available liposomes include transfectace (DDAB/DOPE) and DOTAP/DOPE (Boerhinger). Other cationic liposomes can be prepared from readily available materials using techniques well known in the art. See, *eg.* Szoka (1978) *Proc. Natl. Acad. Sci. USA* 75:4194-4198; WO90/11092 for a description of the synthesis of DOTAP (1,2-bis(oleoyloxy)-3-(trimethylammonio)propane) liposomes.

15 Similarly, anionic and neutral liposomes are readily available, such as from Avanti Polar Lipids (Birmingham, AL), or can be easily prepared using readily available materials. Such materials include phosphatidyl choline, cholesterol, phosphatidyl ethanolamine, dioleoylphosphatidyl choline (DOPC), dioleoylphosphatidyl glycerol (DOPG), dioleoylphoshatidyl ethanolamine (DOPE), among others. These materials can also be mixed with the DOTMA and DOTAP starting materials in appropriate ratios. Methods for making liposomes using these materials are well known in the art.

The liposomes can comprise multilammelar vesicles (MLVs), small unilamellar vesicles (SUVs), or large unilamellar vesicles (LUVs). The various liposome-nucleic acid complexes are prepared using methods known in the art. See eg. Straubinger (1983) Meth. Immunol. 101:512-527; Szoka (1978) Proc. Natl. Acad. Sci. USA 75:4194-4198; Papahadjopoulos (1975) Biochim. Biophys. Acta 394:483; Wilson (1979) Cell 17:77); Deamer & Bangham (1976) Biochim. Biophys. Acta 443:629; Ostro (1977) Biochem. Biophys. Res. Commun. 76:836; Fraley (1979) Proc. Natl. Acad. Sci. USA 76:3348); Enoch & Strittmatter (1979) Proc. Natl. Acad. Sci. USA 76:145; Fraley (1980) J. Biol. Chem. (1980) 255:10431; Szoka & Papahadjopoulos (1978) Proc. Natl. Acad. Sci. USA 75:145; and Schaefer-Ridder (1982) Science 215:166.

#### E.Lipoproteins

5

10

15

20

25

30

In addition, lipoproteins can be included with the polynucleotide/polypeptide to be delivered. Examples of lipoproteins to be utilized include: chylomicrons, HDL, IDL, LDL, and VLDL. Mutants, fragments, or fusions of these proteins can also be used. Also, modifications of naturally occurring lipoproteins can be used, such as acetylated LDL. These lipoproteins can target the delivery of polynucleotides to cells expressing lipoprotein receptors. Preferably, if lipoproteins are including with the polynucleotide to be delivered, no other targeting ligand is included in the composition.

Naturally occurring lipoproteins comprise a lipid and a protein portion. The protein portion are known as apoproteins. At the present, apoproteins A, B, C, D, and E have been isolated and identified. At least two of these contain several proteins, designated by Roman numerals, AI, AII, AIV; CI, CII, CIII.

A lipoprotein can comprise more than one apoprotein. For example, naturally occurring chylomicrons comprises of A, B, C, and E, over time these lipoproteins lose A and acquire C and E apoproteins. VLDL comprises A, B, C, and E apoproteins, LDL comprises apoprotein B; and HDL comprises apoproteins A, C, and E.

The amino acid of these apoproteins are known and are described in, for example, Breslow (1985) Annu Rev. Biochem 54:699; Law (1986) Adv. Exp Med. Biol. 151:162; Chen (1986) J Biol Chem 261:12918; Kane (1980) Proc Natl Acad Sci USA 77:2465; and Utermann (1984) Hum Genet 65:232.

Lipoproteins contain a variety of lipids including, triglycerides, cholesterol (free and esters), and phopholipids. The composition of the lipids varies in naturally occurring lipoproteins. For example, chylomicrons comprise mainly triglycerides. A more detailed description of the lipid content of naturally occurring lipoproteins can be found, for example, in *Meth. Enzymol.* 128 (1986). The composition of the lipids are chosen to aid in conformation of the apoprotein for receptor binding activity. The composition of lipids can also be chosen to facilitate hydrophobic interaction and association with the polynucleotide binding molecule.

Naturally occurring lipoproteins can be isolated from serum by ultracentrifugation, for instance. Such methods are described in *Meth. Enzymol.* (supra); Pitas (1980) J. Biochem. 255:5454-5460 and Mahey (1979) J Clin. Invest 64:743-750. Lipoproteins can also be produced by in vitro or recombinant methods by expression of the apoprotein genes in a desired host cell. See, for example, Atkinson (1986) Annu Rev Biophys Chem 15:403 and Radding (1958) Biochim Biophys Acta 30:

O 99/24578 PCT/IB98/01665

443. Lipoproteins can also be purchased from commercial suppliers, such as Biomedical Techniologies, Inc., Stoughton, Massachusetts, USA. Further description of lipoproteins can be found in Zuckermann et al. PCT/US97/14465.

#### F.Polycationic Agents

10

15

Polycationic agents can be included, with or without lipoprotein, in a composition with the desired polynucleotide/polypeptide to be delivered.

Polycationic agents, typically, exhibit a net positive charge at physiological relevant pH and are capable of neutralizing the electrical charge of nucleic acids to facilitate delivery to a desired location. These agents have both in vitro, ex vivo, and in vivo applications. Polycationic agents can be used to deliver nucleic acids to a living subject either intramuscularly, subcutaneously, etc.

The following are examples of useful polypeptides as polycationic agents: polylysine, polyarginine, polyornithine, and protamine. Other examples include histones, protamines, human serum albumin, DNA binding proteins, non-histone chromosomal proteins, coat proteins from DNA viruses, such as (X174, transcriptional factors also contain domains that bind DNA and therefore may be useful as nucleic aid condensing agents. Briefly, transcriptional factors such as C/CEBP, c-jun, c-fos, AP-1, AP-2, AP-3, CPF, Prot-1, Sp-1, Oct-1, Oct-2, CREP, and TFIID contain basic domains that bind DNA sequences.

Organic polycationic agents include: spermine, spermidine, and purtrescine.

The dimensions and of the physical properties of a polycationic agent can be extrapolated from the list above, to construct other polypeptide polycationic agents or to produce synthetic polycationic agents.

Synthetic polycationic agents which are useful include, for example, DEAE-dextran, polybrene. Lipofectin<sup>TM</sup>, and lipofectAMINE<sup>TM</sup> are monomers that form polycationic complexes when combined with polynucleotides/polypeptides.

## 25 <u>Immunodiagnostic Assays</u>

Neisserial antigens of the invention can be used in immunoassays to detect antibody levels (or, conversely, anti-Neisserial antibodies can be used to detect antigen levels). Immunoassays based on well defined, recombinant antigens can be developed to replace invasive diagnostics methods. Antibodies to Neisserial proteins within biological samples, including for example, blood or serum

samples, can be detected. Design of the immunoassays is subject to a great deal of variation, and a variety of these are known in the art. Protocols for the immunoassay may be based, for example, upon competition, or direct reaction, or sandwich type assays. Protocols may also, for example, use solid supports, or may be by immunoprecipitation. Most assays involve the use of labeled antibody or polypeptide; the labels may be, for example, fluorescent, chemiluminescent, radioactive, or dye molecules. Assays which amplify the signals from the probe are also known; examples of which are assays which utilize biotin and avidin, and enzyme-labeled and mediated immunoassays, such as ELISA assays.

Kits suitable for immunodiagnosis and containing the appropriate labeled reagents are constructed by packaging the appropriate materials, including the compositions of the invention, in suitable containers, along with the remaining reagents and materials (for example, suitable buffers, salt solutions, etc.) required for the conduct of the assay, as well as suitable set of assay instructions.

#### Nucleic Acid Hybridisation

5

10

15

20

25

"Hybridization" refers to the association of two nucleic acid sequences to one another by hydrogen bonding. Typically, one sequence will be fixed to a solid support and the other will be free in solution. Then, the two sequences will be placed in contact with one another under conditions that favor hydrogen bonding. Factors that affect this bonding include: the type and volume of solvent; reaction temperature; time of hybridization; agitation; agents to block the non-specific attachment of the liquid phase sequence to the solid support (Denhardt's reagent or BLOTTO); concentration of the sequences; use of compounds to increase the rate of association of sequences (dextran sulfate or polyethylene glycol); and the stringency of the washing conditions following hybridization. See Sambrook *et al.* [supra] Volume 2, chapter 9, pages 9.47 to 9.57.

"Stringency" refers to conditions in a hybridization reaction that favor association of very similar sequences over sequences that differ. For example, the combination of temperature and salt concentration should be chosen that is approximately 120 to 200°C below the calculated Tm of the hybrid under study. The temperature and salt conditions can often be determined empirically in preliminary experiments in which samples of genomic DNA immobilized on filters are hybridized to the sequence of interest and then washed under conditions of different stringencies. See Sambrook *et al.* at page 9.50.

Variables to consider when performing, for example, a Southern blot are (1) the complexity of the DNA being blotted and (2) the homology between the probe and the sequences being detected. The

10

15

20

25

30

total amount of the fragment(s) to be studied can vary a magnitude of 10, from 0.1 to 1μg for a plasmid or phage digest to 10<sup>-9</sup> to 10<sup>-8</sup> g for a single copy gene in a highly complex eukaryotic genome. For lower complexity polynucleotides, substantially shorter blotting, hybridization, and exposure times, a smaller amount of starting polynucleotides, and lower specific activity of probes can be used. For example, a single-copy yeast gene can be detected with an exposure time of only 1 hour starting with 1 μg of yeast DNA, blotting for two hours, and hybridizing for 4-8 hours with a probe of 10<sup>8</sup> cpm/μg. For a single-copy mammalian gene a conservative approach would start with 10 μg of DNA, blot overnight, and hybridize overnight in the presence of 10% dextran sulfate using a probe of greater than 10<sup>8</sup> cpm/μg, resulting in an exposure time of ~24 hours.

Several factors can affect the melting temperature (Tm) of a DNA-DNA hybrid between the probe and the fragment of interest, and consequently, the appropriate conditions for hybridization and washing. In many cases the probe is not 100% homologous to the fragment. Other commonly encountered variables include the length and total G+C content of the hybridizing sequences and the ionic strength and formamide content of the hybridization buffer. The effects of all of these factors can be approximated by a single equation:

Tm= 
$$81 + 16.6(\log_{10}Ci) + 0.4[\%(G + C)] - 0.6(\%formamide) - 600/n - 1.5(\%mismatch)$$
.

where Ci is the salt concentration (monovalent ions) and n is the length of the hybrid in base pairs (slightly modified from Meinkoth & Wahl (1984) Anal. Biochem. 138: 267-284).

In designing a hybridization experiment, some factors affecting nucleic acid hybridization can be conveniently altered. The temperature of the hybridization and washes and the salt concentration during the washes are the simplest to adjust. As the temperature of the hybridization increases (*ie.* stringency), it becomes less likely for hybridization to occur between strands that are nonhomologous, and as a result, background decreases. If the radiolabeled probe is not completely homologous with the immobilized fragment (as is frequently the case in gene family and interspecies hybridization experiments), the hybridization temperature must be reduced, and background will increase. The temperature of the washes affects the intensity of the hybridizing band and the degree of background in a similar manner. The stringency of the washes is also increased with decreasing salt concentrations.

In general, convenient hybridization temperatures in the presence of 50% formamide are 42°C for a probe with is 95% to 100% homologous to the target fragment, 37°C for 90% to 95% homology.

and 32°C for 85% to 90% homology. For lower homologies, formamide content should be lowered and temperature adjusted accordingly, using the equation above. If the homology between the probe and the target fragment are not known, the simplest approach is to start with both hybridization and wash conditions which are nonstringent. If non-specific bands or high background are observed after autoradiography, the filter can be washed at high stringency and reexposed. If the time required for exposure makes this approach impractical, several hybridization and/or washing stringencies should be tested in parallel.

# Nucleic Acid Probe Assays

5

10

15

20

25

30

Methods such as PCR, branched DNA probe assays, or blotting techniques utilizing nucleic acid probes according to the invention can determine the presence of cDNA or mRNA. A probe is said to "hybridize" with a sequence of the invention if it can form a duplex or double stranded complex, which is stable enough to be detected.

The nucleic acid probes will hybridize to the Neisserial nucleotide sequences of the invention (including both sense and antisense strands). Though many different nucleotide sequences will encode the amino acid sequence, the native Neisserial sequence is preferred because it is the actual sequence present in cells. mRNA represents a coding sequence and so a probe should be complementary to the coding sequence; single-stranded cDNA is complementary to mRNA, and so a cDNA probe should be complementary to the non-coding sequence.

The probe sequence need not be identical to the Neisserial sequence (or its complement) — some variation in the sequence and length can lead to increased assay sensitivity if the nucleic acid probe can form a duplex with target nucleotides, which can be detected. Also, the nucleic acid probe can include additional nucleotides to stabilize the formed duplex. Additional Neisserial sequence may also be helpful as a label to detect the formed duplex. For example, a non-complementary nucleotide sequence may be attached to the 5' end of the probe, with the remainder of the probe sequence being complementary to a Neisserial sequence. Alternatively, non-complementary bases or longer sequences can be interspersed into the probe, provided that the probe sequence has sufficient complementarity with the a Neisserial sequence in order to hybridize therewith and thereby form a duplex which can be detected.

The exact length and sequence of the probe will depend on the hybridization conditions, such as temperature, salt condition and the like. For example, for diagnostic applications, depending on the

25

30

complexity of the analyte sequence, the nucleic acid probe typically contains at least 10-20 nucleotides, preferably 15-25, and more preferably at least 30 nucleotides, although it may be shorter than this. Short primers generally require cooler temperatures to form sufficiently stable hybrid complexes with the template.

Probes may be produced by synthetic procedures, such as the triester method of Matteucci et al. [J. Am. Chem. Soc. (1981) 103:3185], or according to Urdea et al. [Proc. Natl. Acad. Sci. USA (1983) 80: 7461], or using commercially available automated oligonucleotide synthesizers.

The chemical nature of the probe can be selected according to preference. For certain applications, DNA or RNA are appropriate. For other applications, modifications may be incorporated eg. backbone modifications, such as phosphorothioates or methylphosphonates, can be used to increase in vivo half-life, alter RNA affinity, increase nuclease resistance etc. [eg. see Agrawal & Iyer (1995) Curr Opin Biotechnol 6:12-19; Agrawal (1996) TIBTECH 14:376-387]; analogues such as peptide nucleic acids may also be used [eg. see Corey (1997) TIBTECH 15:224-229; Buchardt et al. (1993) TIBTECH 11:384-386].

Alternatively, the polymerase chain reaction (PCR) is another well-known means for detecting small amounts of target nucleic acids. The assay is described in: Mullis *et al.* [*Meth. Enzymol.* (1987) 155: 335-350]; US patents 4,683,195 and 4,683,202. Two "primer" nucleotides hybridize with the target nucleic acids and are used to prime the reaction. The primers can comprise sequence that does not hybridize to the sequence of the amplification target (or its complement) to aid with duplex stability or, for example, to incorporate a convenient restriction site. Typically, such sequence will flank the desired Neisserial sequence.

A thermostable polymerase creates copies of target nucleic acids from the primers using the original target nucleic acids as a template. After a threshold amount of target nucleic acids are generated by the polymerase, they can be detected by more traditional methods, such as Southern blots. When using the Southern blot method, the labelled probe will hybridize to the Neisserial sequence (or its complement).

Also, mRNA or cDNA can be detected by traditional blotting techniques described in Sambrook et al [supra]. mRNA, or cDNA generated from mRNA using a polymerase enzyme, can be purified and separated using gel electrophoresis. The nucleic acids on the gel are then blotted onto a solid support, such as nitrocellulose. The solid support is exposed to a labelled probe and then washed

to remove any unhybridized probe. Next, the duplexes containing the labeled probe are detected. Typically, the probe is labelled with a radioactive moiety.

#### **BRIEF DESCRIPTION OF THE DRAWINGS**

Figures 1-20 show biochemical data obtained in the Examples, and also sequence analysis, for ORFs 37, 5, 2, 15, 22, 28, 32, 4, 61, 76, 89, 97, 106, 138, 23, 25, 27, 79, 85 and 132. M1 and M2 are molecular weight markers. Arrows indicate the position of the main recombinant product or, in Western blots, the position of the main N.meningitidis immunoreactive band. TP indicates N.meningitidis total protein extract; OMV indicates N.meningitidis outer membrane vesicle preparation. In bactericidal assay results: a diamond (♠) shows preimmune data; a triangle (♠) shows GST control data; a circle (♠) shows data with recombinant N.meningitidis protein. Computer analyses show a hydrophilicity plot (upper), an antigenic index plot (middle), and an AMPHI analysis (lower). The AMPHI program has been used to predict T-cell epitopes [Gao et al. (1989) J. Immunol. 143:3007; Roberts et al. (1996) AIDS Res Hum Retrovir 12:593; Quakyi et al. (1992) Scand J Immunol suppl.11:9) and is available in the Protean package of DNASTAR, Inc. (1228 South Park Street, Madison, Wisconsin 53715 USA).

#### **EXAMPLES**

25

The examples describe nucleic acid sequences which have been identified in *N.meningitidis*, along with their putative translation products, and also those of *N.gonorrhoeae*. Not all of the nucleic acid sequences are complete *ie*. they encode less than the full-length wild-type protein.

- 20 The examples are generally in the following format:
  - a nucleotide sequence which has been identified in N.meningitidis (strain B)
  - the putative translation product of this sequence
  - a computer analysis of the translation product based on database comparisons
  - corresponding gene and protein sequences identified in *N.meningitidis* (strain A) and in *N.gonorrhoeae*
  - a description of the characteristics of the proteins which indicates that they might be suitably antigenic
  - results of biochemical analysis (expression, purification, ELISA, FACS etc.)

10

The examples typically include details of sequence identity between species and strains. Proteins that are similar in sequence are generally similar in both structure and function, and the sequence identity often indicates a common evolutionary origin. Comparison with sequences of proteins of known function is widely used as a guide for the assignment of putative protein function to a new sequence and has proved particularly useful in whole-genome analyses.

PCT/IB98/01665

Sequence comparisons were performed at NCBI (http://www.ncbi.nlm.nih.gov) using the algorithms BLAST, BLAST2, BLAST1, BLAST2, tBLAST2, tBLAST2, tBLAST2, & tBLAST2 [eg. see also Altschul et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* 25:2289-3402]. Searches were performed against the following databases: non-redundant GenBank+EMBL+DDBJ+PDB sequences and non-redundant GenBank CDS translations+PDB+SwissProt+SPupdate+PIR sequences.

To compare Meningococcal and Gonococcal sequences, the tBLASTx algorithm was used, as implemented at http://www.genome.ou.edu/gono\_blast.html. The FASTA algorithm was also used to compare the ORFs (from GCG Wisconsin Package, version 9.0).

Dots within nucleotide sequences (eg. position 495 in SEQ ID 11) represent nucleotides which have been arbitrarily introduced in order to maintain a reading frame. In the same way, double-underlined nucleotides were removed. Lower case letters (eg. position 496 in SEQ ID 11) represent ambiguities which arose during alignment of independent sequencing reactions (some of the nucleotide sequences in the examples are derived from combining the results of two or more experiments).

Nucleotide sequences were scanned in all six reading frames to predict the presence of hydrophobic domains using an algorithm based on the statistical studies of Esposti *et al.* [Critical evaluation of the hydropathy of membrane proteins (1990) *Eur J Biochem* 190:207-219]. These domains represent potential transmembrane regions or hydrophobic leader sequences.

Open reading frames were predicted from fragmented nucleotide sequences using the program ORFFINDER (NCBI).

Underlined amino acid sequences indicate possible transmembrane domains or leader sequences in the ORFs, as predicted by the PSORT algorithm (http://www.psort.nibb.ac.jp). Functional domains were also predicted using the MOTIFS program (GCG Wisconsin & PROSITE).

Various tests can be used to assess the *in vivo* immunogencity of the proteins identified in the examples. For example, the proteins can be expressed recombinantly and used to screen patient sera by immunoblot. A positive reaction between the protein and patient serum indicates that the patient has previously mounted an immune response to the protein in question *ie*. the protein is an immunogen. This method can also be used to identify immunodominant proteins.

The recombinant protein can also be conveniently used to prepare antibodies eg. in a mouse. These can be used for direct confirmation that a protein is located on the cell-surface. Labelled antibody (eg. fluorescent labelling for FACS) can be incubated with intact bacteria and the presence of label on the bacterial surface confirms the location of the protein.

In particular, the following methods (A) to (S) were used to express, purify and biochemically characterise the proteins of the invention:

## A) Chromosomal DNA preparation

5

15

20

25

N.meningitidis strain 2996 was grown to exponential phase in 100ml of GC medium, harvested by centrifugation, and resuspended in 5ml buffer (20% Sucrose, 50mM Tris-HCl, 50mM EDTA, pH8). After 10 minutes incubation on ice, the bacteria were lysed by adding 10ml lysis solution (50mM NaCl, 1% Na-Sarkosyl, 50μg/ml Proteinase K), and the suspension was incubated at 37°C for 2 hours. Two phenol extractions (equilibrated to pH 8) and one ChCl<sub>3</sub>/isoamylalcohol (24:1) extraction were performed. DNA was precipitated by addition of 0.3M sodium acetate and 2 volumes ethanol, and was collected by centrifugation. The pellet was washed once with 70% ethanol and redissolved in 4ml buffer (10mM Tris-HCl, 1mM EDTA, pH 8). The DNA concentration was measured by reading the OD at 260 nm.

# B) Oligonucleotide design

Synthetic oligonucleotide primers were designed on the basis of the coding sequence of each ORF, using (a) the meningococcus B sequence when available, or (b) the gonococcus/meningococcus A sequence, adapted to the codon preference usage of meningococcus as necessary. Any predicted signal peptides were omitted, by deducing the 5'-end amplification primer sequence immediately downstream from the predicted leader sequence.

For most ORFs, the 5' primers included two restriction enzyme recognition sites (BamHI-NdeI, BamHI-NheI, or EcoRI-NheI, depending on the gene's own restriction pattern); the 3' primers included

10

15

20

a XhoI restriction site. This procedure was established in order to direct the cloning of each amplification product (corresponding to each ORF) into two different expression systems: pGEX-KG (using either BamHI-XhoI or EcoRI-XhoI), and pET21b+ (using either NdeI-XhoI or NheI-XhoI).

5'-end primer tail: CGCGGATCCCATATG (BamHI-NdeI)

CGCGGATCCGCTAGC (BamHI-NheI)

CCGGAATTCTAGCTAGC (EcoRI-NheI)

3'-end primer tail: CCCGCTCGAG (XhoI)

For ORFs 5, 15, 17, 19, 20, 22, 27, 28, 65 & 89, two different amplifications were performed to clone each ORF in the two expression systems. Two different 5' primers were used for each ORF; the same 3' XhoI primer was used as before:

5'-end primer tail: GGAATTCCATATGGCCATGG (NdeI)

5'-end primer tail: CGGGATCC (BamHI)

ORF 76 was cloned in the pTRC expression vector and expressed as an amino-terminus His-tag fusion. In this particular case, the predicted signal peptide was included in the final product. *NheI-BamHI* restriction sites were incorporated using primers:

5'-end primer tail: GATCAGCTAGCCATATG (NheI)

3'-end primer tail: CGGGATCC (BamHI)

As well as containing the restriction enzyme recognition sequences, the primers included nucleotides which hybridized to the sequence to be amplified. The number of hybridizing nucleotides depended on the melting temperature of the whole primer, and was determined for each primer using the formulae:

 $T_m = 4 (G+C)+ 2 (A+T)$  (tail excluded)

 $T_m = 64.9 + 0.41 \text{ (% GC)} - 600/N$  (whole primer)

The average melting temperature of the selected oligos were 65-70°C for the whole oligo and 50-55°C for the hybridising region alone.

Table I (page 487) shows the forward and reverse primers used for each amplification. In certain cases, it will be noted that the sequence of the primer does not exactly match the sequence in the ORF. When initial amplifications were performed, the complete 5' and/or 3' sequence was not

known for some meningococcal ORFs, although the corresponding sequences had been identified in gonococcus. For amplification, the gonococcal sequences could thus be used as the basis for primer design, altered to take account of codon preference. In particular, the following codons were changed: ATA→ATT; TCG→TCT; CAG→CAA; AAG→AAA; GAG→GAA; CGA→CGC; CGG→CGC; GGG→GGC. Italicised nucleotides in Table I indicate such a change. It will be appreciated that, once the complete sequence has been identified, this approach is generally no longer necessary.

Oligos were synthesized by a Perkin Elmer 394 DNA/RNA Synthesizer, eluted from the columns in 2ml NH<sub>4</sub>OH, and deprotected by 5 hours incubation at 56°C. The oligos were precipitated by addition of 0.3M Na-Acetate and 2 volumes ethanol. The samples were then centrifuged and the pellets resuspended in either 100µl or 1ml of water. OD<sub>260</sub> was determined using a Perkin Elmer Lambda Bio spectophotometer and the concentration was determined and adjusted to 2-10pmol/µl.

# C) Amplification

5

10

15

20

The standard PCR protocol was as follows: 50-200ng of genomic DNA were used as a template in the presence of 20-40µM of each oligo, 400-800µM dNTPs solution, 1x PCR buffer (including 1.5mM MgCl<sub>2</sub>), 2.5 units *TaqI* DNA polymerase (using Perkin-Elmer AmpliTaQ, GIBCO Platinum, Pwo DNA polymerase, or Tahara Shuzo Taq polymerase).

In some cases, PCR was optimsed by the addition of 10µl DMSO or 50µl 2M betaine.

After a hot start (adding the polymerase during a preliminary 3 minute incubation of the whole mix at 95°C), each sample underwent a double-step amplification: the first 5 cycles were performed using as the hybridization temperature the one of the oligos excluding the restriction enzymes tail, followed by 30 cycles performed according to the hybridization temperature of the whole length oligos. The cycles were followed by a final 10 minute extension step at 72°C.

The standard cycles were as follows:

	Denaturation	Hybridisation	Elongation
First 5 cycles	30 seconds	30 seconds	30-60 seconds
	95°C	50-55℃	72°C
Last 30 cycles	30 seconds	30 seconds	30-60 seconds

95°C	65-70°C	72°C
 		L

The elongation time varied according to the length of the ORF to be amplified.

The amplifications were performed using either a 9600 or a 2400 Perkin Elmer GeneAmp PCR System. To check the results, 1/10 of the amplification volume was loaded onto a 1-1.5% agarose gel and the size of each amplified fragment compared with a DNA molecular weight marker.

The amplified DNA was either loaded directly on a 1% agarose gel or first precipitated with ethanol and resuspended in a suitable volume to be loaded on a 1% agarose gel. The DNA fragment corresponding to the right size band was then eluted and purified from gel, using the Qiagen Gel Extraction Kit, following the instructions of the manufacturer. The final volume of the DNA fragment was 30µl or 50µl of either water or 10mM Tris, pH 8.5.

# 10 D) Digestion of PCR fragments

15

20

25

The purified DNA corresponding to the amplified fragment was split into 2 aliquots and double-digested with:

- NdeI/XhoI or NheI/XhoI for cloning into pET-21b+ and further expression of the protein as a C-terminus His-tag fusion
- BamHI/XhoI or EcoRI/XhoI for cloning into pGEX-KG and further expression of the protein as N-terminus GST fusion.
  - For ORF 76, Nhel/BamHI for cloning into pTRC-HisA vector and further expression of the protein as N-terminus His-tag fusion.
- EcoRI/PstI, EcoRI/SalI, SalI/PstI for cloning into pGex-His and further expression of the protein as N-terminus His-tag fusion

Each purified DNA fragment was incubated (37°C for 3 hours to overnight) with 20 units of each restriction enzyme (New England Biolabs ) in a either 30 or 40µl final volume in the presence of the appropriate buffer. The digestion product was then purified using the QIAquick PCR purification kit, following the manufacturer's instructions, and eluted in a final volume of 30 or 50µl of either water or 10mM Tris-HCl, pH 8.5. The final DNA concentration was determined by 1% agarose gel electrophoresis in the presence of titrated molecular weight marker.

# E) Digestion of the cloning vectors (pET22B, pGEX-KG, pTRC-His A, and pGex-His)

10μg plasmid was double-digested with 50 units of each restriction enzyme in 200μl reaction volume in the presence of appropriate buffer by overnight incubation at 37°C. After loading the whole digestion on a 1% agarose gel, the band corresponding to the digested vector was purified from the gel using the Qiagen QIAquick Gel Extraction Kit and the DNA was eluted in 50μl of 10mM Tris-HCl, pH 8.5. The DNA concentration was evaluated by measuring OD<sub>260</sub> of the sample, and adjusted to 50μg/μl. 1μl of plasmid was used for each cloning procedure.

The vector pGEX-His is a modified pGEX-2T vector carrying a region encoding six histidine residues upstream to the thrombin cleavage site and containing the multiple cloning site of the vector pTRC99 (Pharmacia).

## F) Cloning

5

10

15

20

25

The fragments corresponding to each ORF, previously digested and purified, were ligated in both pET22b and pGEX-KG. In a final volume of 20µl, a molar ratio of 3:1 fragment/vector was ligated using 0.5µl of NEB T4 DNA ligase (400 units/µl), in the presence of the buffer supplied by the manufacturer. The reaction was incubated at room temperature for 3 hours. In some experiments, ligation was

performed using the Boheringer "Rapid Ligation Kit", following the manufacturer's instructions.

In order to introduce the recombinant plasmid in a suitable strain, 100µl E. coli DH5 competent cells were incubated with the ligase reaction solution for 40 minutes on ice, then at 37°C for 3 minutes, then, after adding 800µl LB broth, again at 37°C for 20 minutes. The cells were then centrifuged at maximum speed in an Eppendorf microfuge and resuspended in approximately 200µl of the supernatant. The suspension was then plated on LB ampicillin (100mg/ml).

The screening of the recombinant clones was performed by growing 5 randomly-chosen colonies overnight at 37°C in either 2ml (pGEX or pTC clones) or 5ml (pET clones) LB broth + 100µg/ml ampicillin. The cells were then pelletted and the DNA extracted using the Qiagen QIAprep Spin Miniprep Kit, following the manufacturer's instructions, to a final volume of 30µl. 5µl of each individual miniprep (approximately 1g) were digested with either NdeI/XhoI or BamHI/XhoI and the whole digestion loaded onto a 1-1.5% agarose gel (depending on the expected insert size), in parallel with the molecular weight marker (1Kb DNA Ladder, GIBCO). The screening of the positive clones was made on the base of the correct insert size.

For the cloning of ORFs 110, 111, 113, 115, 119, 122, 125 & 130, the double-digested PCR product was ligated into double-digested vector using *EcoRI-PstI* cloning sites or, for ORFs 115 & 127, *EcoRI-SaII* or, for ORF 122, *SaII-PstI*. After cloning, the recombinant plasmids were introduced in the *E.coli* host W3110. Individual clones were grown overnight at 37°C in L-broth with 50µl/ml ampicillin.

-54-

# G) Expression

5

10

15

20

25

30

Each ORF cloned into the expression vector was transformed into the strain suitable for expression of the recombinant protein product. 1μl of each construct was used to transform 30μl of *E.coli* BL21 (pGEX vector), *E.coli* TOP 10 (pTRC vector) or *E.coli* BL21-DE3 (pET vector), as described above. In the case of the pGEX-His vector, the same *E.coli* strain (W3110) was used for initial cloning and expression. Single recombinant colonies were inoculated into 2ml LB+Amp (100μg/ml), incubated at 37°C overnight, then diluted 1:30 in 20ml of LB+Amp (100μg/ml) in 100ml flasks, making sure that the OD<sub>600</sub> ranged between 0.1 and 0.15. The flasks were incubated at 30°C into gyratory water bath shakers until OD indicated exponential growth suitable for induction of expression (0.4-0.8 OD for pET and pTRC vectors; 0.8-1 OD for pGEX and pGEX-His vectors). For the pET, pTRC and pGEX-His vectors, the protein expression was induced by addition of 1mM IPTG, whereas in the case of pGEX system the final concentration of IPTG was 0.2mM. After 3 hours incubation at 30°C, the final concentration of the sample was checked by OD. In order to check expression, 1ml of each sample was removed, centrifuged in a microfuge, the pellet resuspended in PBS, and analysed by 12% SDS-PAGE with Coomassie Blue staining. The whole sample was centrifuged at 6000g and the pellet resuspended in PBS for further use.

# H) GST-fusion proteins large-scale purification.

A single colony was grown overnight at 37°C on LB+Amp agar plate. The bacteria were inoculated into 20ml of LB+Amp liquid colture in a water bath shaker and grown overnight. Bacteria were diluted 1:30 into 600ml of fresh medium and allowed to grow at the optimal temperature (20-37°C) to OD<sub>550</sub> 0.8-1. Protein expression was induced with 0.2mM IPTG followed by three hours incubation. The culture was centrifuged at 8000rpm at 4°C. The supernatant was discarded and the bacterial pellet was resuspended in 7.5ml cold PBS. The cells were disrupted by sonication on ice for 30 sec at 40W using a Branson sonifier B-15, frozen and thawed twice and centrifuged again. The supernatant was collected and mixed with 150µl Glutatione-Sepharose 4B resin (Pharmacia)

10

15

20

(previously washed with PBS) and incubated at room temperature for 30 minutes. The sample was centrifuged at 700g for 5 minutes at 4°C. The resin was washed twice with 10ml cold PBS for 10 minutes, resuspended in 1ml cold PBS, and loaded on a disposable column. The resin was washed twice with 2ml cold PBS until the flow-through reached OD<sub>280</sub> of 0.02-0.06. The GST-fusion protein was eluted by addition of 700μl cold Glutathione elution buffer (10mM reduced glutathione, 50mM Tris-HCl) and fractions collected until the OD<sub>280</sub> was 0.1. 21μl of each fraction were loaded on a 12% SDS gel using either Biorad SDS-PAGE Molecular weight standard broad range (M1) (200, 116.25, 97.4, 66.2, 45, 31, 21.5, 14.4, 6.5 kDa) or Amersham Rainbow Marker (M2) (220, 66, 46, 30, 21.5, 14.3 kDa) as standards. As the MW of GST is 26kDa, this value must be added to the MW of each GST-fusion protein.

#### I) His-fusion solubility analysis (ORFs 111-129)

To analyse the solubility of the His-fusion expression products, pellets of 3ml cultures were resuspended in buffer M1 [500μl PBS pH 7.2]. 25μl lysozyme (10mg/ml) was added and the bacteria were incubated for 15 min at 4°C. The pellets were sonicated for 30 sec at 40W using a Branson sonifier B-15, frozen and thawed twice and then separated again into pellet and supernatant by a centrifugation step. The supernatant was collected and the pellet was resuspended in buffer M2 [8M urea, 0.5M NaCl, 20mM imidazole and 0.1M NaH<sub>2</sub> PO<sub>4</sub>] and incubated for 3 to 4 hours at 4°C. After centrifugation, the supernatant was collected and the pellet was resuspended in buffer M3 [6M guanidinium-HCl, 0.5M NaCl, 20mM imidazole and 0.1M NaH<sub>2</sub>PO<sub>4</sub>] overnight at 4°C. The supernatants from all steps were analysed by SDS-PAGE.

The proteins expressed from ORFs 113, 119 and 120 were found to be soluble in PBS, whereas ORFs 111, 122, 126 and 129 need urea and ORFs 125 and 127 need guanidium-HCl for their solubilization.

# J) His-fusion large-scale purification.

A single colony was grown overnight at 37°C on a LB + Amp agar plate. The bacteria were inoculated into 20ml of LB+Amp liquid culture and incubated overnight in a water bath shaker. Bacteria were diluted 1:30 into 600ml fresh medium and allowed to grow at the optimal temperature (20-37°C) to OD<sub>550</sub> 0.6-0.8. Protein expression was induced by addition of 1mM IPTG and the culture further incubated for three hours. The culture was centrifuged at 8000rpm at 4°C, the supernatant was discarded and the bacterial pellet was resuspended in 7.5ml of either (i) cold

buffer A (300mM NaCl, 50mM phosphate buffer, 10mM imidazole, pH 8) for soluble proteins or (ii) buffer B (urea 8M, 10mM Tris-HCl, 100mM phosphate buffer, pH 8.8) for insoluble proteins.

The cells were disrupted by sonication on ice for 30 sec at 40W using a Branson sonifier B-15, frozen and thawed two times and centrifuged again.

For insoluble proteins, the supernatant was stored at -20°C, while the pellets were resuspended in 2ml 5 buffer C (6M guanidine hydrochloride, 100mM phosphate buffer, 10mM Tris-HCl, pH 7.5) and treated in a homogenizer for 10 cycles. The product was centrifuged at 13000rpm for 40 minutes.

Supernatants were collected and mixed with 150µl Ni2+-resin (Pharmacia) (previously washed with either buffer A or buffer B, as appropriate) and incubated at room temperature with gentle agitation for 30 minutes. The sample was centrifuged at 700g for 5 minutes at 4°C. The resin was washed twice with 10ml buffer A or B for 10 minutes, resuspended in 1ml buffer A or B and loaded on a disposable column. The resin was washed at either (i) 4°C with 2ml cold buffer A or (ii) room temperature with 2ml buffer B, until the flow-through reached OD<sub>280</sub> of 0.02-0.06.

The resin was washed with either (i) 2ml cold 20mM imidazole buffer (300mM NaCl, 50mM phosphate buffer, 20mM imidazole, pH 8) or (ii) buffer D (urea 8M, 10mM Tris-HCl, 100mM phosphate buffer, pH 6.3) until the flow-through reached the O.D<sub>280</sub> of 0.02-0.06. The His-fusion protein was eluted by addition of 700µl of either (i) cold elution buffer A (300mM NaCl, 50mM phosphate buffer, 250mM imidazole, pH 8) or (ii) elution buffer B (urea 8M, 10mM Tris-HCl, 100mM phosphate buffer, pH 4.5) and fractions collected until the O.D<sub>280</sub> was 0.1. 21µl of each fraction were loaded on a 12% SDS gel.

# K) His-fusion proteins renaturation

10

15

20

25

10% glycerol was added to the denatured proteins. The proteins were then diluted to 20µg/ml using dialysis buffer I (10% glycerol, 0.5M arginine, 50mM phosphate buffer, 5mM reduced glutathione, 0.5mM oxidised glutathione, 2M urea, pH 8.8) and dialysed against the same buffer at 4°C for 12-14 hours. The protein was further dialysed against dialysis buffer II (10% glycerol, 0.5M arginine, 50mM phosphate buffer, 5mM reduced glutathione, 0.5mM oxidised glutathione, pH 8.8) for 12-14 hours at 4°C. Protein concentration was evaluated using the formula:

Protein (mg/ml) = 
$$(1.55 \times OD_{280}) - (0.76 \times OD_{260})$$

15

20

25

30

# L) His-fusion large-scale purification (ORFs 111-129)

500ml of bacterial cultures were induced and the fusion proteins were obtained soluble in buffer M1, M2 or M3 using the procedure described above. The crude extract of the bacteria was loaded onto a Ni-NTA superflow column (Quiagen) equilibrated with buffer M1, M2 or M3 depending on the solubilization buffer of the fusion proteins. Unbound material was eluted by washing the column with the same buffer. The specific protein was eluted with the corresponding buffer containing 500mM imidazole and dialysed against the corresponding buffer without imidazole. After each run the columns were sanitized by washing with at least two column volumes of 0.5 M sodium hydroxide and reequilibrated before the next use.

#### 10 M) Mice immunisations

20μg of each purified protein were used to immunise mice intraperitoneally. In the case of ORFs 2, 4, 15, 22, 27, 28, 37, 76, 89 and 97, Balb-C mice were immunised with Al(OH)<sub>3</sub> as adjuvant on days 1, 21 and 42, and immune response was monitored in samples taken on day 56. For ORFs 44, 106 and 132, CD1 mice were immunised using the same protocol. For ORFs 25 and 40, CD1 mice were immunised using Freund's adjuvant, rather than AL(OH)<sub>3</sub>, and the same immunisation protocol was used, except that the immune response was measured on day 42, rather than 56. Similarly, for ORFs 23, 32, 38 and 79, CD1 mice were immunised with Freund's adjuvant, but the immune response was measured on day 49.

#### N) ELISA assay (sera analysis)

The acapsulated MenB M7 strain was plated on chocolate agar plates and incubated overnight at 37°C. Bacterial colonies were collected from the agar plates using a sterile dracon swab and inoculated into 7ml of Mueller-Hinton Broth (Difco) containing 0.25% Glucose. Bacterial growth was monitored every 30 minutes by following OD<sub>620</sub>. The bacteria were let to grow until the OD reached the value of 0.3-0.4. The culture was centrifuged for 10 minutes at 10000rpm. The supernatant was discarded and bacteria were washed once with PBS, resuspended in PBS containing 0.025% formaldehyde, and incubated for 2 hours at room temperature and then overnight at 4°C with stirring. 100µl bacterial cells were added to each well of a 96 well Greiner plate and incubated overnight at 4°C. The wells were then washed three times with PBT washing buffer (0.1% Tween-20 in PBS). 200µl of saturation buffer (2.7% Polyvinylpyrrolidone 10 in water) was added to each well and the plates incubated for 2 hours at 37°C. Wells were washed

15

20

25

three times with PBT. 200µl of diluted sera (Dilution buffer: 1% BSA, 0.1% Tween-20, 0.1% NaN<sub>3</sub> in PBS) were added to each well and the plates incubated for 90 minutes at 37°C. Wells were washed three times with PBT. 100µl of HRP-conjugated rabbit anti-mouse (Dako) serum diluted 1:2000 in dilution buffer were added to each well and the plates were incubated for 90 minutes at 37°C. Wells were washed three times with PBT buffer. 100µl of substrate buffer for HRP (25ml of citrate buffer pH5, 10mg of O-phenildiamine and 10µl of H<sub>2</sub>O) were added to each well and the plates were left at room temperature for 20 minutes. 100µl H<sub>2</sub>SO<sub>4</sub> was added to each well and OD<sub>490</sub> was followed. The ELISA was considered positive when OD<sub>490</sub> was 2.5 times the respective pre-immune sera.

# 10 O) FACScan bacteria Binding Assay procedure.

The acapsulated MenB M7 strain was plated on chocolate agar plates and incubated overnight at 37°C. Bacterial colonies were collected from the agar plates using a sterile dracon swab and inoculated into 4 tubes containing 8ml each Mueller-Hinton Broth (Difco) containing 0.25% glucose. Bacterial growth was monitored every 30 minutes by following OD<sub>620</sub>. The bacteria were let to grow until the OD reached the value of 0.35-0.5. The culture was centrifuged for 10 minutes at 4000rpm. The supernatant was discarded and the pellet was resuspended in blocking buffer (1% BSA, 0.4% NaN<sub>3</sub>) and centrifuged for 5 minutes at 4000rpm. Cells were resuspended in blocking buffer to reach OD620 of 0.07. 100µl bacterial cells were added to each well of a Costar 96 well plate. 100µl of diluted (1:200) sera (in blocking buffer) were added to each well and plates incubated for 2 hours at 4°C. Cells were centrifuged for 5 minutes at 4000rpm, the supernatant aspirated and cells washed by addition of 200µl/well of blocking buffer in each well. 100µl of R-Phicoerytrin conjugated F(ab)<sub>2</sub> goat anti-mouse, diluted 1:100, was added to each well and plates incubated for 1 hour at 4°C. Cells were spun down by centrifugation at 4000rpm for 5 minutes and washed by addition of 200µl/well of blocking buffer. The supernatant was aspirated and cells resuspended in  $200\mu l$ /well of PBS, 0.25% formaldehyde. Samples were transferred to FACScan tubes and read. The condition for FACScan setting were: FL1 on, FL2 and FL3 off; FSC-H threshold:92; FSC PMT Voltage: E 02; SSC PMT: 474; Amp. Gains 7.1; FL-2 PMT: 539; compensation values: 0.

## P) OMV preparations

5

10

20

25

Bacteria were grown overnight on 5 GC plates, harvested with a loop and resuspended in 10 ml 20mM Tris-HCl. Heat inactivation was performed at 56°C for 30 minutes and the bacteria disrupted by sonication for 10 minutes on ice (50% duty cycle, 50% output). Unbroken cells were removed by centrifugation at 5000g for 10 minutes and the total cell envelope fraction recovered by centrifugation at 50000g at 4°C for 75 minutes. To extract cytoplasmic membrane proteins from the crude outer membranes, the whole fraction was resuspended in 2% sarkosyl (Sigma) and incubated at room temperature for 20 minutes. The suspension was centrifuged at 10000g for 10 minutes to remove aggregates, and the supernatant further ultracentrifuged at 50000g for 75 minutes to pellet the outer membranes. The outer membranes were resuspended in 10mM Tris-HCl, pH8 and the protein concentration measured by the Bio-Rad Protein assay, using BSA as a standard.

# Q) Whole Extracts preparation

Bacteria were grown overnight on a GC plate, harvested with a loop and resuspended in 1ml of 20mM Tris-HCl. Heat inactivation was performed at 56°C for 30 minutes.

# 15 R) Western blotting

Purified proteins (500ng/lane), outer membrane vesicles (5μg) and total cell extracts (25μg) derived from MenB strain 2996 were loaded on 15% SDS-PAGE and transferred to a nitrocellulose membrane. The transfer was performed for 2 hours at 150mA at 4°C, in transferring buffer (0.3 % Tris base, 1.44 % glycine, 20% methanol). The membrane was saturated by overnight incubation at 4°C in saturation buffer (10% skimmed milk, 0.1% Triton X100 in PBS). The membrane was washed twice with washing buffer (3% skimmed milk, 0.1% Triton X100 in PBS) and incubated for 2 hours at 37°C with mice sera diluted 1:200 in washing buffer. The membrane was washed twice and incubated for 90 minutes with a 1:2000 dilution of horseradish peroxidase labelled antimouse Ig. The membrane was washed twice with 0.1% Triton X100 in PBS and developed with the Opti-4CN Substrate Kit (Bio-Rad). The reaction was stopped by adding water.

#### S) Bactericidal assay

MC58 strain was grown overnight at 37°C on chocolate agar plates. 5-7 colonies were collected and used to inoculate 7ml Mueller-Hinton broth. The suspension was incubated at 37°C on a nutator and let to grow until OD<sub>620</sub> was 0.5-0.8. The culture was aliquoted into sterile 1.5ml Eppendorf

-60-

tubes and centrifuged for 20 minutes at maximum speed in a microfuge. The pellet was washed once in Gey's buffer (Gibco) and resuspended in the same buffer to an  $OD_{620}$  of 0.5, diluted 1:20000 in Gey's buffer and stored at 25°C.

50μl of Gey's buffer/1% BSA was added to each well of a 96-well tissue culture plate. 25μl of diluted mice sera (1:100 in Gey's buffer/0.2% BSA) were added to each well and the plate incubated at 4°C. 25μl of the previously described bacterial suspension were added to each well. 25μl of either heat-inactivated (56°C waterbath for 30 minutes) or normal baby rabbit complement were added to each well. Immediately after the addition of the baby rabbit complement, 22μl of each sample/well were plated on Mueller-Hinton agar plates (time 0). The 96-well plate was incubated for 1 hour at 37°C with rotation and then 22μl of each sample/well were plated on Mueller-Hinton agar plates (time 1). After overnight incubation the colonies corresponding to time 0 and time 1 hour were counted.

Table II (page 493) gives a summary of the cloning, expression and prurification results.

# Example 1

5

10

15 The following partial DNA sequence was identified in N.meningitidis <SEQ ID 1>:

```
1 ATGAAACAGA CAGTCAA.AT GCTTGCCGCC GCCCTGATTG CCTTGGGCTT
                    GAACCGACCG GTGTGGNCGG ATGACGTATC GGATTTTCGG GAAAACTTGC
                    A.GCGGCAGC ACAGGGAAAT GCAGCAGCCC AATACAATTT GGGCGCAATG
               101
                    TAT. TACAAA GGACGCGCGT GCGCCGGGAT GATGCTGAAG CGGTCAGATG
               151
                    GTATCGGCAG CCGGCGGAAC AGGGGTTAGC CCAAGCCCAA TACAATTTGG
                    GCTGGATGTA TGCCAACGGG CGCGC.GTGC GCCAAGATGA TACCGAAGCG
20
               201
                251
                    GTCAGATGGT ATCGGCAGGC GGCAGCGCAG GGGGTTGTCC AAGCCCAATA
                     CAATTTGGGC GTGATATATG CCGAAGGACG TGGAGTGCGC CAAGACGATG
                301
                351
                     TCGAAGCGGT CAGATGGTTT CGGCAGGCGG CAGCGCAGGG GGTAGCCCAA
                401
                     GCCCAAAACA ATTTGGGCGT GATGTATGCC GAAAGANCGC GCGTGCGCCA
25
                451
                     AGACCG...
```

This corresponds to the amino acid sequence <SEQ ID 2; ORF37>:

```
1 MKQTVXMLAA ALIALGLNRP VWXDDVSDFR ENLXAAAQGN AAAQYNLGAM
51 YXQRTRVRRD DAEAVRWYRQ PAEQGLAQAQ YNLGWMYANG RXVRQDDTEA
30 101 VRWYRQAAAQ GVVQAQYNLG VIYAEGRGVR QDDVEAVRWF RQAAAQGVAQ
151 AQNNLGVMYA ERXRVRQD...
```

Further work revealed the complete nucleotide sequence <SEQ ID 3>:

```
ATGAAACAGA CAGTCAAATG GCTTGCCGCC GCCCTGATTG CCTTGGGCTT
                    GAACCGAGCG GTGTGGGCGG ATGACGTATC GGATTTTCGG GAAAACTTGC
                51
                    AGGCGGCAGC ACAGGGAAAT GCAGCAGCCC AATACAATTT GGGCGCAATG
               101
                    TATTACAAAG GACGCGGCGT GCGCCGGGAT GATGCTGAAG CGGTCAGATG
35
                    GTATCGGCAG GCGGCGGAAC AGGGGTTAGC CCAAGCCCAA TACAATTTGG
               151
               201
                    GCTGGATGTA TGCCAACGGG CGCGGCGTGC GCCAAGATGA TACCGAAGCG
                    GTCAGATGGT ATCGGCAGGC GGCAGCGCAG GGGGTTGTCC AAGCCCAATA
                251
                301
                    CAATTTGGGC GTGATATATG CCGAAGGACG TGGAGTGCGC CAAGACGATG
                351
                    TCGAAGCGGT CAGATGGTTT CGGCAGGCGG CAGCGCAGGG GGTAGCCCAA
40
                401
                    GCCCAAAACA ATTTGGGCGT GATGTATGCC GAAAGACGCG GCGTGCGCCA
                451
                    AGACCGCGCC CTTGCACAAG AATGGTTTGG CAAGGCTTGT CAAAACGGAG
                501
                    ACCAAGACGG CTGCGACAAT GACCAACGCC TGAAGGCGGG TTATTGA
                551
```

This corresponds to the amino acid sequence <SEQ ID 4; ORF37-1>:

```
1 MKQTVKWLAA ALIALGLNRA VWADDVSDFR ENLQAAAQGN AAAQYNLGAM
51 YYKGRGVRRD DAEAVRWYRQ AAEQGLAQAQ YNLGWMYANG RGVRQDDTEA
101 VRWYRQAAAQ GVVQAQYNLG VIYAEGRGVR QDDVEAVRWF RQAAAQGVAQ
5 151 AQNNLGVMYA ERRGVRQDRA LAQEWFGKAC QNGDQDGCDN DQRLKAGY*
```

Further work identified the corresponding gene in strain A of N. meningitidis <SEQ ID 5>:

```
1 ATGAAACAGA CAGTCAAATG GCTTGCCGCC GCCCTGATTG CCTTGGGCTT
51 GAACCAAGCG GTGTGGGCGG ATGACGTATC GGATTTTCGG GAAAACTTGC
101 AGGCGGCAGC ACAGGGAAAT GCAGCAGCCC AAAACAATTT GGGCGTGATG
151 TATGCCGAAA GACGCGGCGT GCGCCAAGAC CGCGCCCTTG CACAAGAATG
201 GCTTGGCAAG GCTTGTCAAA ACGGATACCA AGACAGCTGC GACAATGACC
251 AACGCCTGAA AGCGGGTTAT TGA
```

This encodes a protein having amino acid sequence <SEQ ID 6; ORF37a>:

```
1 MKQTVKWLAA ALIALGLNQA VWADDVSDFR ENLQAAAQGN AAAQNNLGVM
51 YAERRGVRQD RALAQEWLGK ACQNGYQDSC DNDQRLKAGY *
```

The originally-identified partial strain B sequence (ORF37) shows 68.0% identity over a 75aa overlap with ORF37a:

```
20
                                                                50
                     MKQTVXMLAAALIALGLNRPVWXDDVSDFRENLXAAAQGNAAAQYNLGAMYXQRTRVRRD
20
         orf37.pep
                           MKQTVKWLAAALIALGLNQAVWADDVSDFRENLQAAAQGNAAAQNNLGVMYAERRGVRQD
         orf37a
                            10
                                     20
                                              30
                                                       40
                                                                50
25
                            70
                                     80
                                              90
                                                      100
                     DAEAVRWYRQPAEQGLAQAQYNLGWMYANGRXVRQDDTEAVRWYRQAAAQGVVQAQYNLG
         orf37.pep
                     RALAQEWLGKACQNGYQDSCDNDQRLKAGYX
         orf37a
                            70
                                     80
```

30 Further work identified the corresponding gene in N. gonorrhoeae <SEQ ID 7 >:

This encodes a protein having amino acid sequence <SEQ ID 8; ORF37ng>:

40 1 MKQTVKWLAA ALIALGLNQA VWAGDVSDFR ENLQAAEQGN AAAQFNLGVM
51 YENGQGVRQD YVQAVQWYRK ASEQGDAQAQ YNLGLMYYDG RGVRQDLALA
101 QQWLGKACQN GDQNSCDNDQ RLKAGY\*

The originally-identified partial strain B sequence (ORF37) shows 64.9% identity over a 111aa overlap with ORF37ng:

45	orf37.pep	MKQTVXMLAAALIALGLNRPVWXDDVSDFRENLXAAAQGNAAAQYNLGAMYXQRTRVRRD	60
	orf37ng	MKQTVKWLAAALIALGLNQAVWAGDVSDFRENLQAAEQGNAAAQFNLGVMYENGQGVRQD	60
50	orf37.pep	DAEAVRWYRQPAEQGLAQAQYNLGWMYANGRXVRQDDTEAVRWYRQAAAQGVVQAQYNLG ::  :  :	120
	orf37ng	YVQAVQWYRKASEQGDAQAQYNLGLMYYDGRGVRQDLALAQQWLGKACQNGDQNSCDNDQ	120
	orf37.pep	VIYAEGRGVRQDDVEAVRWFRQAAAQGVAQAQNNLGVMYAERXRVRQD 168	
55	orf37ng	RLKAGY 126	

The complete strain B sequence (ORF37-1) and ORF37ng show 51.5% identity in 198 aa overlap:

	orf37-1.pep	10 20 30 40 50 60 MKQTVKWLAAALIALGLNRAVWADDVSDFRENLQAAAQGNAAAQYNLGAMYYKGRGVRRD
5	orf37ng	MKQTVKWLAAALIALGLNQAVWAGDVSDFRENLQAAEQGNAAAQFNLGVMYENGQGVRQD 10 20 30 40 50 60
	orf37-1.pep	70 80 90 100 110 120 DAEAVRWYRQAAEQGLAQAQYNLGWMYANGRGVRQDDTEAVRWYRQAAAQGVVQAQYNLG
10	orf37ng	::  :  : :
		130 140 150 160 170 180 VIYAEGRGVRQDDVEAVRWFRQAAAQGVAQAQNNLGVMYAERRGVRQDRALAQEWFGKAC
15	orf37-1.pep orf37ng	: :     
	01137.119	100
20	orf37-1.pep	190 199 QNGDQDGCDNDQRLKAGYX      ::
	orf37ng	QNGDQNSCDNDQRLKAGYX 110 120

Computer analysis of these amino acid sequences indicates a putative leader sequence, and it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF37-1 (11kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 1A shows the results of affinity purification of the GST-fusion protein, and Figure 1B shows the results of expression of the His-fusion in *E.coli*. Purified GST-fusion protein was used to immunise mice, whose sera were used for ELISA (positive result), FACS analysis (Figure 1C), and a bactericidal assay (Figure 1D). These experiments confirm that ORF37-1 is a surface-exposed protein, and that it is a useful immunogen.

35 Figure 1E shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF37-1.

# Example 2

30

The following partial DNA sequence was identified in N.meningitidis <SEQ ID 9>:

```
TTCGGCGA CATCGGCGGT TTGAAGGTCA ATGCCCCCGT CAAATCCGCA
GGCGTATTGG TCGGGCGCGT CGGCGCTATC GGACTTGACC CGAAATCCTA
TCAGGCGAGG GTGCGCCTCG ATTTGGACGG CAAGTATCAG TTCAGCAGCG
ACGTTTCCGC GCAAATCCTG ACTTCsGGAC TTTTGGGCGA GCAGTACATC
GGGCTGCAGC AGGGCGGCGA CACGGAAAAC CTTGCTGCCG GCGACACCAT
CTCCGTAACC AGTTCTGCAA TGGTTCTGGA AAACCTTATC GGCAAATTCA
TGACGAGTTT TGCCGAGAAA AATGCCGACG GCGCAATGC GGAAAAAGCC

45
GCCGAATAA
```

This corresponds to the amino acid sequence <SEQ ID 10>:

<sup>1</sup> FGDIGGLKVN APVKSAGVLV GRVGAIGLDP KSYQARVRLD LDGKYQFSSD 51 VSAQILTSGL LGEQYIGLQQ GGDTENLAAG DTISVTSSAM VLENLIGKFM

101 TSFAEKNADG GNAEKAAE\*

Computer analysis of this amino acid sequence gave the following results:

Homology with a hypothetical *H.influenzae* protein (ybrd.haein; accession number p45029) SEO ID 9 and ybrd.haein show 48.4% as identity in 122 as overlap:

```
5
         yrbd.h LGIGALVFLGLRVANVQGFAETKSYTVTATFDNIGGLKVRAPLKIGGVVIGRVSAITLDE
                                          N.m
                                          FGDIGGLKVNAPVKSAGVLVGRVGAIGLDP
                                                 10
                                                          20
                                                                   30
10
                 80
                          90
                                  100
                                           110
                                                    120
                                                             130
         yrbd.h KSYLPKVSIAINQEYNEIPENSSLSIKTSGLLGEQYIALTMGFDDGDTAMLKNGSQIQDT
                111: 1:1: 1
         N.m
                KSYQARVRLDLDGKY-QFSSDVSAQILTSGLLGEQYIGLQQG
                                                        GDTENLAAGDTISVT
15
                                50
                                                  70
                                         60
                                                              80
                 140
                         150
                                    160
         yrbd.h TSAMVLEDLIGQFL--YGSKKSDGNEKSESTEQ
                :||||||:|||:|: :::|::||:: ::::|:
20
         N.m
                SSAMVLENLIGKFMTSFAEKNADGGNAEKAAEX
                         100
                                  110
```

#### Homology with a predicted ORF from N.gonorrhoeae

SEQ ID 9 shows 99.2% identity over a 118aa overlap with a predicted ORF from N. gonorrhoeae:

25		20	30	40	50	60	70	
	yrbd				VYADFGDIGGI			GLDP
	N.m	•					 SVLVGRVGAI	
	M · In				rgbiggi	10	20	30
30								
		80	90	100	110	120	130	
	yrbd	KSYQARVI	RLDLDGKYQE	SSDVSAQIL	TSGLLGEQYIC	LQQGGDTEN	LAAGDTISVT	SSAM
					11111111111			1111
	N.m	KSYQARVI	RLDLDGKYQE	SSDVSAQIL	TSGLLGEQYIC	LQQGGDTEN	LAAGDTISVT	SSAM
35	,		40	50	60	70	80	90
		140	150	160				
	yrbd	VLENLIGH	KEMTSFAEKN	NAEGGNAEKA	AEX			
	3	1111111		11:111111	111			
40	N.m	VLENLIG	KEMTSFAEKN	NADGGNAEKA	AEX			
		. 1	L00	110	120			

The complete yrbd *H.influenzae* sequence has a leader sequence and it is expected that the full-length homologous *N.meningitidis* protein will also have one. This suggests that it is either a membrane protein, a secreted protein, or a surface protein and that the protein, or one of its epitopes, could be a useful antigen for vaccines or diagnostics.

#### Example 3

45

The following partial DNA sequence was identified in N. meningitidis <SEQ ID 11>:

	1	ATTTTGATAT	ACCTCATCCG	CAAGAATCTA	GGTTCGCCCG	TCTTCTTCTT
•	51	TCAGGAACGC	CCCGGAAAGG	ACGGAAAACC	TTTTAAAATG	GTCAAATTCC
50	101	GTTCCATGCG	CGACGGCTTG	TATTCAGACG	GCATTCCGCT	GCCCGACGGA
•	151	GAACGCCTGA	CACCGTTCGG	CAAAAAACTG	CGTGCCGcCA	GTWTGGACGA
	201	ACTGCCTGAA	TTATGGAATA	TCTTAAAAGG	CGAGATGAGC	CTGGTCGGCC
	251	CCCGCCCGCT	GCTGATGCAA	TATCTGCCGC	TGTACGACAA	CTTCCAAAAC
	301	CGCCGCCACG	AAATGAAACC	CGGCATTACC	GGCTGGGCGC	AGGTCAACGG

```
GCGCAACGCg CTTTCGTGGG ACGAAAAATT CGCCTGCGAT GTTTGGTATA
               351
                      TCGACCACTT CAGCCTGTGC CTCGACATCA AAATCCTACT GCTGACGGTT
               401
                       AAAAAGTAT TAATCAAGGA AGGGATTTCC GCACAGGGCG AACA.aCCAT
               451
                       GCCCCCTTTC ACAGGAAAAC GCAAACTCGC CGTCGTCGGT GCGGGCGGAC
               501
                       ACGGAAAAGT CGTTGCCGAC CTTGCCGCCG CACTCGGCCG GTACAGGGAA
5
               551
                       ATCGTTTTC TGGACGACCG CGCACAAGGC AGCGTCAACG GCTTTTCCGT
                601
                       CATCGGCACG ACGCTGCTGC TTGAAAACAG TTTATCGCCC GAACAATACG
                651
                       ACGTCGCCGT CGCCGTCGGC AACAACCGCA TCCGCCGCCA AATCGCCGAA
                701
                       AAAGCCGCCG CGCTCGGCTT CGCCCTGCCC GTACTGGTTC ATCCGGACGC
                751
                       GACCGTCTCG CCTTCTGCAA CAGTCGGACA AGGCAGCGTC GTTATGGCGA
10
                801
                       AAGCGGTCG..
                851
     This corresponds to the amino acid sequence <SEQ ID 12; ORF3>:
                     ..ILIYLIRKNL GSPVFFFQER PGKDGKPFKM VKFRSMRDGL YSDGIPLPDG
                       ERLTPFGKKL RAASXDELPE LWNILKGEMS LVGPRPLLMQ YLPLYDNFQN
                 51
                       RRHEMKPGIT GWAQVNGRNA LSWDEKFACD VWYIDHFSLC LDIKILLLTV
                101
15
                       KKVLIKEGIS AQGEXTMPPF TGKRKLAVVG AGGHGKVVAD LAAALGRYRE
                151
                       IVFLDDRAQG SVNGFSVIGT TLLLENSLSP EQYDVAVAVG NNRIRRQIAE
                201
                       KAAALGFALP VLVHPDATVS PSATVGQGSV VMAKAV..
                251
      Further sequence analysis revealed the complete nucleotide sequence <SEQ ID 13>:
                     ATGAGTAAAT TCTTCAAACG CCTGTTTGAC ATTGTTGCCT CCGCCTCGGG
20
                     ACTGATTTTC CTCTCGCCAG TATTTTTGAT TTTGATATAC CTCATCCGCA
                      AGAATCTAGG TTCGCCCGTC TTCTTCTTTC AGGAACGCCC CGGAAAGGAC
                 101
                      GGAAAACCTT TTAAAATGGT CAAATTCCGT TCCATGCGCG ACGCGCTTGA
                151
                      TTCAGACGGC ATTCCGCTGC CCGACGGAGA ACGCCTGACA CCGTTCGGCA
                 201
                      AAAAACTGCG TGCCGCCAGT TTGGACGAAC TGCCTGAATT ATGGAATATC
                 251
25
                      TTAAAAGGCG AGATGAGCCT GGTCGGCCCC CGCCCGCTGC TGATGCAATA
                      TCTGCCGCTG TACGACAACT TCCAAAACCG CCGCCACGAA ATGAAACCCG
                 351
                      GCATTACCGG CTGGGCGCAG GTCAACGGGC GCAACGCGCT TTCGTGGGAC
                 401
                      GAAAAATTCG CCTGCGATGT TTGGTATATC GACCACTTCA GCCTGTGCCT
                 451
                      CGACATCAAA ATCCTACTGC TGACGGTTAA AAAAGTATTA ATCAAGGAAG
                 501
 30
                      GGATTTCCGC ACAGGGCGAA GCCACCATGC CCCCTTTCAC AGGAAAACGC
                 551
                     AAACTCGCCG TCGTCGGTGC GGGCGGACAC GGAAAAGTCG TTGCCGACCT
                 601
                      TGCCGCCGCA CTCGGCCGGT ACAGGGAAAT CGTTTTTCTG GACGACCGCG
                 651
                      CACAAGGCAG CGTCAACGGC TTTTCCGTCA TCGGCACGAC GCTGCTGCTT
                 701
                      GAAAACAGTT TATCGCCCGA ACAATACGAC GTCGCCGTCG CCGTCGGCAA
 35
                 751
                      CAACCGCATC CGCCGCCAAA TCGCCGAAAA AGCCGCCGCG CTCGGCTTCG
                 801
                      CCCTGCCCGT TCTGGTTCAT CCGGACGCGA CCGTCTCGCC TTCTGCAACA
                 851
                      GTCGGACAAG GCAGCGTCGT TATGGCGAAA GCCGTCGTAC AGGCAGGCAG
                 901
                      CGTATTGAAA GACGGCGTGA TTGTGAACAC TGCCGCCACC GTCGATCACG
                 951
                      ACTGCCTGCT TAACGCTTTC GTCCACATCA GCCCAGGCGC GCACCTGTCG
                1001
 40
                      GGCAACACGC ATATCGGCGA AGAAAGCTGG ATAGGCACGG GCGCGTGCAG
                 1051
                      CCGCCAGCAG ATCCGTATCG GCAGCCGCGC AACCATTGGA GCGGGCGCAG
                 1101
                      TCGTCGTACG CGACGTTTCA GACGGCATGA CCGTCGCGGG CAATCCGGCA
                 1151
                      AAGCCGCTGC CGCGCAAAAA CCCCGAGACC TCGACAGCAT AA
       This corresponds to the amino acid sequence <SEQ ID 14; ORF3-1>:
  45
                      MSKFFKRLFD IVASASGLIF LSPVFLILIY LIRKNLGSPV FFFQERPGKD
                       GKPFKMVKFR SMRDALDSDG IPLPDGERLT PFGKKLRAAS LDELPELWNI
                      LKGEMSLVGP RPLLMQYLPL YDNFQNRRHE MKPGITGWAQ VNGRNALSWD
                      EKFACDVWYI DHFSLCLDIK ILLLTVKKVL IKEGISAQGE ATMPPFTGKR
                  201 KLAVVGAGGH GKVVADLAAA LGRYREIVFL DDRAQGSVNG FSVIGTTLLL
  50
                       ENSLSPEQYD VAVAVGNNRI RRQIAEKAAA LGFALPVLVH PDATVSPSAT
                  251
                       VGQGSVVMAK AVVQAGSVLK DGVIVNTAAT VDHDCLLNAF VHISPGAHLS
                  301
                       GNTHIGEESW IGTGACSRQQ IRIGSRATIG AGAVVVRDVS DGMTVAGNPA
                  351
                  401 KPLPRKNPET STA*
```

55 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from N. meningitidis (strain A)

ORF3 shows 93.0% identity over a 286aa overlap with an ORF (ORF3a) from strain A of N. meningitidis:

		•			
	orf3.pep	•	ILI	<u>(LI</u> RKNLGSPVFE	FFQERPGKDGKPFKMVKFR
	orf3a	MSKFFKRLFDIVA 10	ASASGLIFLSPVFLILI) 20 30		FFQERPGKDGKPFKMVKFR 50 60
5		10	20 30	, 40	
,		40	50 60	70	80 90
	orf3.pep				<b>GEMSLVGPRPLLMQYLPL</b>
10	orf3a				KGDMSLVGPRPLLMQYLPL
10	,	. 70	80 90	100	110 120
		100	110 120	130	140 150
	orf3.pep				:FSLCLDIKILLLTVKKVL
	02201F0F	11111111111		11:1111:1111	1
15	orf3a				ifs <u>lcldikillltvkk</u> vl
	•	130	140 150	160	170 180
		160	170 180	190	200 210
	orf3.pep				ERYREIVFLDDRAOGSVNG
20	OII3.PCP				
	orf3a			HGKVVAELAAALO	GTYGEIVFLDDRVQGSVNG
		190	200 21	220	230 240
		000	020 040	250	
25	orf3.pep	220	230 240	250 TDDOTAFKAAAT	260 270 GFALPVLVHPDATVSPSAT
23	orrs.pep				
	orf3a				GFALPVLIHPDSTVSPSAT
		250	260 27	0 280	290 300
••					
30	62	280 VGQGSVVMAKAV			
	orf3.pep	VGQGSVVMAKAV			
	orf3a		VQADSVLKDGVIVNTAA	TVDHDCLLDAFV	HISPGAHLSGNTRIGEESW
•		310	320 33		350 360
25	The commiste les	noth ODE2a nuale	eotide sequence <si< th=""><th>70 ID 15&gt; ic.</th><th></th></si<>	70 ID 15> ic.	
35	The complete ici	igui Okt-3a ilucie	conde sequence \Si	3Q 1D 13~ 13.	
	1	ልጥርልርጥልልልጥ ጥርጥጥ	CAAACG CCTGTTTGAC	ATTGTTGCCT	CCGCCTCGGG
•			CGCCAG TATTTTTGAT		
	101	AGAATCTGGG TTCG	CCCGTC TTCTTCTTTC	AGGAACGCCC	CGGAAAGGAC
40	151		AATGGT CAAATTCCGT		
40	201		TGCTGC CCGACGGAGA		
	251 301		GCCAGT TTGGACGAAC GAGCCT GGTCGGCCCC		
	351		ACAACT TCCAAAACCG		
	401		GCGCAG GTCAACGGGC		
45	451		CGACAT CTGGTATATC		
	501		TACTGC TGACGGTTAA		
	551 601		GGCGAA GCCACCATGC		
	651		GCACAT ACGGCGAAAT		
50	701		AACGGC TTCCCCGTCA		
	751	GAAAACAGTT TATC	GCCCGA ACAATTCGAC	ATCGCCGTCG	CCGTCGGCAA
	801		GCCAAA TCGCCGAAAA		
	851		ATTCAT CCGGACTCGA		
55	901 951		CGTCGT TATGGCGAAA GCGTAA TTGTGAACAC		
))	1001		GCTTTC GTCCACATCA		
	1051		CGGCGA AGAAAGCTGG		
	1101		GTATCG GCAGCCGCGC		
	1151		GTTTCA GACGGCATGA		
60	1201	AAACCATTGG CAGG	CAAAAA TACCGAGACC	CTGCGGTCGT	AA
	This is predicted	i to encode a prote	ein having amino ac	id sequence <	SEO ID 16>:
	P				(
	1	MSKFFKRLFD IVAS	SASGLIF LSPVFLILIY	LIRKNLGSPV	FFFQERPGKD
	51	GKPFKMVKFR SMH	ALDSDG ILLPDGERLT	PFGKKLRAAS	LDELPELWNV
			MQYLPL YDNFQNRRHE	_	
65	151		SLCLDIK ILLLTVKKVI		
	201		VAELAAA LGTYGEIVFI NYGNNRI RRQIAEKAAA		
	251	FUSTSEEGED TWAN	MAGMMET VENTURE	TOLWTLATIL	EDSIASESMI

301 VGQGGVVMAK AVVQADSVLK DGVIVNTAAT VDHDCLLDAF VHISPGAHLS 351 GNTRIGEESW IGTGACSRQQ IRIGSRATIG AGAVVVRDVS DGMTVAGNPA 401 KPLAGKNTET LRS\*

Two transmembrane domains are underlined.

ORF3-1 shows 94.6% identity in 410 aa overlap with ORF3a: 5

2	OKL2-1 shows > 110 /1	J 100121115	•				
		10	20	30	40	50	60
		MOVERNET EDINAS	ACCUTELSPV	FI.TI.TYLTRK	NLGSPVFFFQ	ERPGKDGKPFI	MVKFR
	orf3a.pep			1111111111	1111111111		11111
		MSKFFKRLFDIVAS	ASGLTFLSPV	FLILIYLIRK	NLGSPVFFFQ	ERPGKDGKPFI	CLIA L/L L/
	orf3-1	10	20	30	40	50	60
10		10		,			
		70	80	90	100	110	120
	62	CMUDAT DEDCTT.I.E	POGERTAPEGE	KLRAASLDEI	PELWNVLKGD	MSLVGPRPLLI	MQYLPL
	orf3a.pep		. 1 . 1 . 1 <b>. 1 . 1 . 1</b>		11111:11:	. 1	11111
		SMRDALDSDGIPLE	PDGERLTPFGK	KLRAASLDEI	,PELWNILKGE	MSLVGPRPLL	WÕAThr
15	orf3-1	70	80	90	100	110	120
		130	140	150	160	170	180
	orf3a.pep	YDNFQNRRHEMKPO	GITGWAQVNG	RNALSWDERF	ACDIWYIDHFS	PCTDIKITT	TVKKVL
20	Ollyg.beb			111111111	1     2   1   1   1   1	1111111111	11111
20	orf3-1	YDNFQNRRHEMKP	GITGWAQVNGI	RNALSWDEKF	ACDAMAIDHE	さいくいわエピエアロアロ	TAVVAD
	0113 1	130	140	150	160	170	180
				_		230	240
		190	200	210	220		
25	orf3a.pep	IKEGISAQGEATM	PPFTGKRKLA	VVGAGGHGKV	VALLAAALGT	IGETALPODKA	LIIIII
23			111111111		:		CCSVNG
	orf3-1	IKEGISAQGEATM	PPFTGKRKLA	VVGAGGHGKV	VADLAAALGK 220	230	240
		190	200	210	,220	230	
		252	260	270	280	290	300
30		250 FPVIGTTLLLENS	20U	ATTCNIND TODO	TAFKAAAI.GF		TVSPSAT
	orf3a.pep	FPVIGTTLLENS	PSECEDIA	WAGIAINT TUVÕ	IIIIIIIIIII	11111:111:	111111
		FSVIGTTLLENS		AVGNNRTRRO	TAEKAAALGE	ALPVLVHPDAT	TVSPSAT
	orf3-1	250	260	270	280	290	300
		230	200	2.0			
35		310	320	330	340	350	360
	50	TICOCCUMNKAIN.	ADSVIKDGV	TVNTAATVDE	IDCLLDAFVHI	SPGAHLSGNT	RIGEESW
	orf3a.pep			1111111111	1111:111	1111111111	.
	orf3-1	VGQGSVVMAKAVV	/OAGSVLKDGV	IVNTAATVDI	IDCTTNWLAHT	SPGAHLSGNT	HIGEEOM
40	0112-1	310	320	330	340	350	360
40		•=-					
		370	380	390	400	410	
	orf3a.pep	TGTGACSROOIR	igsratigag <i>i</i>	AVVVRDVSDG1	atvagnpakpi	AGKNTETLRS	X
	Ollia.pep						
45	orf3-1	IGTGACSRQQIR:	igsratigag/	AVVVRDVSDGI	MTVAGNPAKPI	LPKKNPLISTA	X
43	0113 1	370	380	390	400	410	
	· ·						

Homology with hypothetical protein encoded by vvfc gene (accession Z71928) of B. subtilis ORF3 and YVFC proteins show 55% aa identity in 170 aa overlap (BLASTp):

50	ORF3	_	IYLIRKNLGSPVFFFQERPGKDGKPFKMVKFRSMRDGLYSDGIPLPDGERLTPFGKKLRA 62 I ++R +GSPVFF Q RPG GKPF + KFR+M D S G LPD RLT G+ +R
	yvfc	27	I THE TOSPYTT O KIND THE TOSPYTT OF TOSPYTT OF THE
55	ORF3	_	ASXDELPELWNILKGEMSLVGPRPLLMQYLPLYDNFQNRRHEMKPGITGWAQVNGRNALS 122 S DELP+L N+LKG++SLVGPRPLLM YLPLY Q RRHE+KPGITGWAQ+NGRNA+S
	yvfc	87	LSIDELPQLLNVLKGDLSLVGPRPLLMDYLPLYTEKQARRHEVKPGITGWAQINGRNAIS 146
	ORF3	_	WDEKFACDVWYIDHFSLCLDXXXXXXXXXXXXXXXEGISAQGEXTMPPFTG 172 W++KF DVWY+D++S LD EGI T FTG
60	yvfc	147	WEKKFELDVWYVDNWSFFLDLKILCLTVRKVLVSEGIQQTNHVTAERFTG 196

# Homology with a predicted ORF from N.gonorrhoeae

ORF3 shows 86.3% identity over a 286aa overlap with a predicted ORF (ORF3.ng) from N. gonorrhoeae:

•	orf3	ILIYLIRKNLGSPVFFFQERPGKDGKPFKMVKFR	. 34
5	orf3ng	MSKAVKRLFDIIASA <u>SGLIVLSPVFLVLIYLI</u> RKNKGSPVFFIRERPGKDGKPFKMVKFR	60
	orf3	SMRDGLYSDGIPLPDGERLTPFGKKLRAASXDELPELWNILKGEMSLVGPRPLLMQYLPL	94
10	orf3ng	:	120
	orf3	YDNFQNRRHEMKPGITGWAQVNGRNALSWDEKFACDVWYIDHFSLCLDIKILLLTVKKVL	154
	orf3ng	::  }	180
15	orf3	IKEGISAQGEXTMPPFTGKRKLAVVGAGGHGKVVADLAAALGRYREIVFLDDRAQGSVNG	214
	orf3ng		240
20	orf3	FSVIGTTLLLENSLSPEQYDVAVAVGNNRIRRQIAEKAAALGFALPVLVHPDATVSPSAT	274
	orf3ng		300
	orf3	VGQGSVVMAKAV	286
25	orf3ng	:             IGQGSVVMAKAVVQAGSVLKDGVIVNTAATVDHDCLLDAFVHISPGAHLSGNTRIGEESR	360

The complete length ORF3ng nucleotide sequence <SEQ ID 17> is:

	1	ATGAGTAAAG	CCGTCAAACG	CCTGTTCGAC	ATCATCGCAT	CCGCATCGGG
	51	GCTGATTGTC	CTGTCGCCCG	TGTTTTTGGT	TTTAATATAC	CTCATCCGCA
30	101	AAAACTTAGG	TTCGCCCGTC	TTCTTCattC	GGGAACGCCc	cgGAAAGGAc
	151	ggaaaacCTT	TTAAAATGGT	CAAATTCCGT	TCCAtgcgcg	acgcgcttGA
	201	TTCAGACGGC	ATTCCGCTGC	CCGATAGCGA	ACGCCTGACC	GATTTCGGCA
	251	AAAAATTACG	CGCCACCAGT	TTGGACGAAC	TTCCTGAATT	ATGGAATGTC
	301	CTCAAAGGCG	AGATGAGCCT	GGTCGGCCCC	CGCCCGCTTT	TGATGCAGTA
35	351	TCTGCCGCTT	TACAACAAAT	TTCAAAACCG	CCGCCACGAA	ATGAAACCGG
	401	GCATTACCGG	CTGGGCGCAG	GTCAACGGGC	GCAACGCGCT	TTCGTGGGAC
	451	GAAAAGTTCT	CCTGCGATGT	TTGGTACACC	GACAATTTCA	GCTTTTGGCT
	501	GGATATGAAA	ATCCTGTTTC	TGACAGTCAA	AAAAGTCTTG	ATTAAAGAAG
	551	GCATTTCGGC	GCAAGGGGAA	GCCACCATGC	CCCCTTTCGC	GGGGAATCGC
40	601	AAACTCGCCG	TTATCGGCGC	GGGCGGACAC	GGCAAAGTCG	TTGCCGAGCT
	651	TGCCGCCGCA	CTCGGCACAT	ACGGCGAAAT	CGTTTTTCTG	GACGACCGCA
	701	CCCAAGGCAG	CGTCAACGGC	TTCCCCGTCA	TCGGCACGAC	GCTGCTGCTT
	751	GAAAACAGTT	TATCGCCCGA	ACAATTCGAC	ATCACCGTCG	CCGTCGGCAA
	801	CAACCGCATC	CGCCGCCAAA	TCACCGAAAA	CGCCGCCGCG	CTCGGCTTCA
45	851	AACTGCCCGT	TCTGATTCAT	CCCGACGCGA	CCGTCTCGCC	TTCTGCAATA
	901	ATCGGACAAG	GCAGCGTCGT	AATGGCGAAA	GCCGTCGTAC	AGGCCGGCAG
	951	CGTATTGAAA	GACGGCGTGA	TTGTGAACAC	TGCCGCCACC	GTCGATCACG
	1001	ACTGCCTGCT	TGACGCTTTC	GtccaCATCA	GCCCGGGCGC	GCACCTGTCG
	1051	GGCAACACGC	GTATCGGCGA	AGAAAGCCGG	ATAGGCACGG	GCGCGTGCAG
50	1101	CCGCCAGCAG	ACAACCGTCG	GCAGCGGGGT	TACCgccgGT	GCAGGGgcGG
	1151	TTATCGTATG	CGACATCCCG	GACGGCATGA	CCGTCGCGGG	CAACCCGGCA
	1201	AAGCCCCTTA	CGGGCAAAAA	CCCCAAGACC	GGGACGGCAT	AA

# This encodes a protein having amino acid sequence <SEQ ID 18>:

	. 1	MSKAVKRLFD	IIASASGLIV	LSPVFLVLIY	LIRKNLGSPV	FFIRERPGKD
55	51	GKPFKMVKFR	SMRDALDSDG	IPLPDSERLT	DFGKKLRATS	LDELPELWNV
	101	LKGEMSLVGP	RPLLMQYLPL	YNKFQNRRHE	MKPGITGWAQ	VNGRNALSWD
	151	EKFSCDVWYT	DNFSFWLDMK	ILFLTVKKVL	IKEGISAQGE	ATMPPFAGNR
	201	KLAVIGAGGH	GKVVAELAAA	LGTYGEIVFL	DDRTQGSVNG	FPVIGTTLLL
	251	ENSLSPEQFD	ITVAVGNNRI	RRQITENAAA	LGFKLPVLIH	PDATVSPSAI
60	301	IGQGSVVMAK	AVVQAGSVLK	DGVIVNTAAT	VDHDCLLDAF	VHISPGAHLS
	351	GNTRIGEESR	IGTGACSRQQ	TTVGSGVTAG	AGAVIVCDIP	DGMTVAGNPA
	401	KPLTGKNPKT	GTA*			

This protein shows 86.9% identity in 413 aa overlap with ORF3-1:

	F						
5	orf3-1.pep	10 MSKFFKRLFDIVASAS                 MSKAVKRLFDIIASAS 10	111 1111	11:111111		1	MVKFR 60
10	orf3-1.pep	70 SMRDALDSDGIPLPDG                SMRDALDSDGIPLPDG 70	1111 111	1111:1111			[[]]]
15	orf3-1.pep	130 YDNFQNRRHEMKPGI  ::          YNKFQNRRHEMKPGI  130	111111111	1111111111:	11111 1:11	:   :   :	11111
20	orf3-1.pep	190 IKEGISAQGEATMPP             IKEGISAQGEATMPP	: :      FAGNRKLAV	IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	:       AELAAALGTY	: GEIVFLDDRT	 QGSVNG
25	orf3-1.pep	190 250 FSVIGTTLLLENSLS	111:1::11		:1:111111	1111:1111	1111
30	orf3ng	FPVIGTTLLLENSLS 250 310	PEQFDITVF 260 320	AVGNNRIRRQI 270 330	TENAAALGFK 280 340	EPVLIHPDAT 290 350	300 360
35	orf3-1.pep orf3ng	VGQGSVVMAKAVVQA :              IGQGSVVMAKAVVQA 310	GSVLKDGV:	CVNTAATVDHI	CLLNAFVHIS	.	11111
40	orf3-1.pep orf3ng	370 IGTGACSRQQIRIGS             :   IGTGACSRQQTTVGS 370	:1 1111	1:1 1: 1111			
	In addition, ORF3ng	g shows significant	homolog	y with a hyp	othetical p	rotein from	B.subtilis:
45	<pre>gnl PID e238668 (271928) hypothetical protein [Bacillus subtilis] &gt;gi 1945702 gnl PID e313004 (294043) hypothetical protein [Bacillus subtilis] &gt;gi 2635938 gnl PID e1186113 (299121) similar to capsular polysaccharide biosynthesis [Bacillus subtilis]Length = 202 Score = 235 bits (594), Expect = 3e-61</pre>						
50	Identities = 114/195 (58%), Positives = 142/195 (72%)  ONEY: 5 VKRLFDIIASASGLIVLSPVFLVLIYLIRKNLGSPVFFIRERPGKDGKPFKMVKFRSMRD 64						MRD 64
	#1	KRLFD+ A+ L S KRLFDLTAAIFLLCCTS	+ L I +	+R +GSPVF	F + RPG G	KPF + KFR+1	MI D
55	*****	LDSDGIPLPDSERLTDF DS G LPD RLT RDSKGNLLPDEVRLTKT	G+ +R S+	DELP+L NVL	KG++SLVGPR	PLLM YLPLY	+
60	Query: 125 Q	NRRHEMKPGITGWAQVN RRHE+KPGITGWAQ+N ARRHEVKPGITGWAQIN	GRNALSWDE	KFSCDVWYTD	NFSFWLDMKI N+SF+LD+KI	LFLTVKKVLI L LTV+KVL+	KEG 184 EG
65	Ī	SAQGEATMPPFAGN 19 T F G+ QQTNHVTAERFTGS 19					

The hypothetical product of yvfc gene shows similarity to EXOY of R.meliloti, an exopolysaccharide production protein. Based on this and on the two predicted transmembrane regions in the homologous N.gonorrhoeae sequence, it is predicted that these proteins, or their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### 5 Example 4

The following partial DNA sequence was identified in N. meningitidis <SEQ ID 19>:

```
.. AACCATATGG CGATTGTCAT CGACGAATAC GGCGGCACAT CCGGCTTGGT
                51
                      CACCTTTGAA GACATCATCG AGCAAATCGT CGGCGAAATC GAAGACGAGT
                101
                      TTGACGAAGA CGATAGCGCC GACAATATCC ATGCCGTTTC TTCAGACACG
                      TGGCGCATCC ATGCAGCTAC CGAAATCGAA GACATCAACA CCTTCTTCGG
10
                151
                      CACGGAATAC AGCATCGAAG AAGCCGACAC CATT.GGCGG CCTGGTCATT
                201
                      CAAGAGTTGG GACATCTGCC CGTGCGCGGC GAAAAAGTCC TTATCGGCGG
                251
                301
                      TTTGCAGTTC ACCGTCGCAC GCGCCGACAA CCGCCGCCTG CATACGCTGA
                351
                      TGGCGACCCG CGTGAAGTAA GC.....ACCGC CGTTTCTGCA
15
                401
                      CAGTTTAG
```

This corresponds to amino acid sequence <SEQ ID 20; ORF5>:

```
1 ..NHMAIVIDEY GGTSGLVTFE DIIEQIVGEI EDEFDEDDSA DNIHAVSSDT
51 WRIHAATEIE DINTFFGTEY SIEEADTIXR PGHSRVGTSA RARRKSPYRR
101 FAVHRRTRQ PPPAYADGDP REVS....XR RFCTV*
```

## 20 Further sequence analysis revealed the complete DNA sequence to be <SEQ ID 21>:

```
ATGGACGGCG CACAACCGAA AACGAATTTT TTTGAACGCC TGATTGCCCG
               51
                   ACTCGCCCGC GAACCCGATT CCGCCGAAGA CGTATTAAAC CTGCTTCGGC
              101
                   AGGCGCACGA GCAGGAAGTT TTTGATGCGG ATACGCTTTT AAGATTGGAA
              151
                   25
              201
                   CAGCCGTATG AACGTTTTAA AAGAAAACGA CAGCATCGAG CGCATCACCG
              251 CCTACGTTAT CGATACCGCC CATTCGCGCT TCCCCGTCAT CGGCGAAGAC
              301 AAAGACGAAG TTTTGGGCAT TTTGCACGCC AAAGACCTGC TCAAATATAT
                   GTTTAACCCC GAGCAGTTCC ACCTCAAATC CATTCTCCGC CCCGCCGTCT
              351
              401
                   TCGTCCCCGA AGGCAAATCG CTGACCGCCC TTTTAAAAGA GTTCCGCGAA
                   CAGCGCAACC ATATGGCGAT TGTCATCGAC GAATACGGCG GCACATCCGG
30
              451
              501
                   CTTGGTCACC TTTGAAGACA TCATCGAGCA AATCGTCGGC GAAATCGAAG
                   ACGAGTTTGA CGAAGACGAT AGCGCCGACA ATATCCATGC CGTTTCTTCC
              551
               601
                   GAACGCTGGC GCATCCATGC AGCTACCGAA ATCGAAGACA TCAACACCTT
                   CTTCGGCACG GAATACAGCA GCGAAGAAGC CGACACCATT CGGCCTGGTC
              651
35
              701
                   ATTCAAGAGT TGGGACATCT GCCCGTGCGC GGCGAAAAAG TCCTTATCGG
                   CGGTTTGCAG TTCACCGTCG CACGCGCCGA CAACCGCCGC CTGCATACGC
              751
              801
                   TGATGGCGAC CCGCGTGAAG TAAGCACCGC CGTTTCTGCA CAGTTTAGGA
                   TGACGGTACG GGCGTTTTCT GTTTCAATCC GCCCCATCCG CCAAACATAA
```

This corresponds to amino acid sequence <SEQ ID 22; ORF5-1>:

```
40

1 MDGAQPKTNF FERLIARLAR EPDSAEDVLN LLRQAHEQEV FDADTLLRLE
51 KVLDFSDLEV RDAMITRSRM NVLKENDSIE RITAYVIDTA HSRFPVIGED
101 KDEVLGILHA KDLLKYMFNP EQFHLKSILR PAVFVPEGKS LTALLKEFRE
151 QRNHMAIVID EYGGTSGLVT FEDIIEQIVG EIEDEFDEDD SADNIHAVSS
201 ERWRIHAATE IEDINTFFGT EYSSEEADTI RPGHSRVGTS ARARRKSPYR
45
251 RFAVHRRTRR QPPPAYADGD PREVSTAVSA QFRMTVRAFS VSIRPIRQT*
```

Further work identified the corresponding gene in strain A of N.meningitidis <SEQ ID 23 >:

5	451 CA 501 TT 551 AT 601 GA 651 TT 701 GT 751 CG 801 CG 851 GG	GTCCCGA AGGCAAATCG CTGACCGCC TTTTAAAAGA GTTCCGCGAA GCGCAACC ATATGGCAAT CGTCATCGAC GAATACGGCG GCACGTCGGG GGGTAACT TTTGAAGACA TCATCGAGCA AATCGTCGGC GACATCGAAG GAGTTTGA CGAAGACGAA AGCGCGGACA ACATCCACGC CGTTTCCGCC ACGCTGGC GCATCCACGC GGCTACCGAA ATCGAAGACA TCAACGCCTT CCGGCACG GAATACAGCA GCGAAGAAGC CGACACCATC GGCGGCCNTG CATTCAGG AATTGGNACA CCTGCCCGTG CGCGCGAAA AAGTCNTTAT GCGNNTTG CANTTCACNG TCGCCNGCGC NGACAACCGC CGCCTGCATA CTGATGGC GACCCGCTG AAGTAAGCTC CGCCGTTTCT GTACAGTTTA ATGACGGT ACGGCGTTT TCTGTTTCAA TCCGCCCCAT CCGCCCANACA						
	901 TA	ein having amino acid sequence <seq 24;="" id="" orf5a="">:</seq>						
15	1 MD 51 KV 101 KD 151 QR 201 ER 251 RF 301 *	GAQPKTNF XXRLIARLAR EPDSAEDVLT LLRQAHEQEV FDADTLLRLE LDFSDLEV RDAMITRSRM NVLKENDSIE RITAYVIDTA HSRFPVIGED EVLGILHA KDLLKYMFNP EQFHLKSILR PAVFVPEGKS LTALLKEFRE NHMAIVID EYGGTSGLVT FEDIIEQIVG DIEDEFDEDE SADNIHAVSA WRIHAATE IEDINAFFGT EYSSEEADTI GGXGHSGIGT PARARRKSXY XXXHXRXR XQPPPAYADG DPREVSSAVS VQFRMTVRAF SVSIRPIRXT						
20	The originally-identified partial strain B sequence (ORF5) shows 54.7% identity over a 124aa							
	overlap with ORF5a:							
25	orf5.pep orf5a	10 20 30  NHMAIVIDEYGGTSGLVTFEDIIEQIVGEI						
30	orf5.pep orf5a	40 50 60 70 80 90  EDEFDEDDSADNIHAVSSDTWRIHAATEIEDINTFFGTEYSIEEADTIXRPGHSRVGTSA        :       ::						
35	orf5.pep orf5a	100 110 120 130  RARRKSPYRRFAVHRRTRRQPPPAYADGDPREVSXXXXXRRFCTV						
	The complete strain B sequence (ORF5-1) and ORF5a show 92.7% identity in 300 aa overlap:							
40	orf5a.pep orf5-1	10 20 30 40 50 60  MDGAQPKTNFXXRLIARLAREPDSAEDVLTLLRQAHEQEVFDADTLLRLEKVLDFSDLEV						
45	orf5a.pep	70 80 90 100 110 120 RDAMITRSRMNVLKENDSIERITAYVIDTAHSRFPVIGEDKDEVLGILHAKDLLKYMFNP						
50	<b></b>	70 80 90 100 110 120  130 140 150 160 170 180  EQFHLKSILRPAVFVPEGKSLTALLKEFREQRNHMAIVIDEYGGTSGLVTFEDIIEQIVG						
55	orf5a.pep							
60	orf5a.pep	190 200 210 220 230 240 DIEDEFDEDESADNIHAVSAERWRIHAATEIEDINAFFGTEYSSEEADTIGGXGHSGIGT :						

```
PARARRKSXYRRXAXHXRXRXQPPPAYADGDPREVSSAVSVQFRMTVRAFSVSIRPIRXT
                         orf5-1
                        {\tt SARARRKSPYRRFAVHRRTRRQPPPAYADGDPREVSTAVSAQFRMTVRAFSVSIRPIRQT
                      240
                                250
                                          260
                                                    270
                                                              280
     Further work identified the a partial DNA sequence in N. gonorrhoeae <SEQ ID 25> which encodes
     a protein having amino acid sequence <SEQ ID 26; ORF5ng>:
                    MDGAQPKTNF FERLIARLAR EPDSAEDVLN LLRQAHEQEV FDADTLTRLE
                    KVLDFAELEV RDAMITRSRM NVLKENDSIE RITAYVIDTA HSRFPVIGED
                51
                    KDEVLGILHA KDLLKYMFNP EQFHLKSVLR PAVFVPEGKS LTALLKEFRE
               101
10
               151
                    QRNHMAIVID EYGGTSGLVT FEDIIEQIVG DIEDEFDEDE SADDIHSVSA
               201
                    ERWRIHAATE IEDINAFFGT EYGSEEADTI RRLGHSGIGT PARARRKSPY
                    RRFAVHRRPR RQPPPAHADG DPREVSRACP HRRFCTV*
     Further analysis revealed the complete gonococcal nucleotide sequence <SEQ ID 27> to be:
                    ATGGACGGCG CACAACCGAA AACAAATTTT TTTGAACGCC TGATTGCCCG
15
                    ACTCGCCCGC GAACCCGATT CCGCCGAAGA CGTATTAAAC CTGCTTCGGC
                    AGGCGCACGA ACAGGAAGTT TTTGATGCCG ACACACTGAC CCGGCTGGAA
               101
                    AAAGTATTGG ACTTTGCCGA GCTGGAAGTG CGCGATGCGA TGATTACGCG
               151
                    CAGCCGCATG AACGTATTGA AAGAAAACGA CAGCATCGAA CGCATCACCG
               201
                    CCTACGTCAT CGATACCGCC CATTCGCGCT TCCCCGTCAT CGGCGAAGAC AAAGACCAAG TTTTGGGCAT TTTGCACGCC AAAGACCTGC TCAAATATAT
               251
20
                301
                    GTTCAACCCC GAGCAGTTCC ACCTGAAATC CGTCTTGCGC CCTGCCGTTT
                351
               401
                    TCGTGCCCGA AGGCAAATCT TTGACCGCCC TTTTAAAAGA GTTCCGCGAA
                    CAGCGCAACC ATATGGCAAT CGTCATCGAC GAATACGGCG GCACGTCGGG
                451
                    TTTGGTCACC TTTGAAGACA TCATCGAGCA AATCGTCGGT GACATCGAAG
               501
25
                    ACGAGTTTGA CGAAGACGAA AGCGccgacg acatCCACTC cgTTTccgCC
                551
                601
                    GAACGCTGGC GCATCCacgc ggctaCCGAA ATCGAAGaca TCAACGCCTT
                    TTTCGGTACG GAatacggca gcgaagaagc cgacaccatc cggcggctTGGTCATTCAGG AATTGGGACA CCTGCCCGTG CGCGGCGAAA AAGTCCTTAt
                651
                701
                    cggcgGTTTG Cagttcaccg tCGCCCGCGC CGACAACCGC CGCCTGCACA
                751
30
                    CGCTGATGGC GACCCGCGTG AAGTAAGCAG AGCCTGCCcg AccgccgttT
                801
                851
                     CTGCacAGTT TAGGAtgACG gtaCGGTCGT TTTCTGTTTC AATCCGCCCC
                901
                    ATCCGCCAAA CATAA
     This encodes a protein having amino acid sequence <SEQ ID 28; ORF5ng-1>:
                    MDGAQPKTNF FERLIARLAR EPDSAEDVLN LLRQAHEQEV FDADTLTRLE
35
                    KVLDFAELEV RDAMITRSRM NVLKENDSIE RITAYVIDTA HSRFPVIGED
                 51
                    KDEVLGILHA KDLLKYMFNP EQFHLKSVLR PAVFVPEGKS LTALLKEFRE
                    QRNHMAIVID EYGGTSGLVT FEDIIEQIVG DIEDEFDEDE SADDIHSVSA
                151
                    ERWRIHAATE IEDINAFFGT EYGSEEADTI RRLGHSGIGT PARARRKSPY
                201
                    RRFAVHRRPR RQPPPAHADG DPREVSRACP TAVSAQFRMT VRSFSVSIRP
                251
40
                     IRQT*
     The originally-identified partial strain B sequence (ORF5) shows 83.1% identity over a 135aa
     overlap with the partial gonococcal sequence (ORF5ng):
           orf5
                                                      NHMAIVIDEYGGTSGLVTFEDIIEQIVGEI
                                                                                       30
                                                      1111111111111111111111111111111111111
45
                        FHLKSVLRPAVFVPEGKSLTALLKEFREQRNHMAIVIDEYGGTSGLVTFEDIIEQIVGDI
           orf5ng
                                                                                      182
                        EDEFDEDDSADNIHAVSSDTWRIHAATEIEDINTFFGTEYSIEEADTIXRPGHSRVGTSA
           orf5
                                                                                       90
                        EDEFDEDESADDIHSVSAERWRIHAATEIEDINAFFGTEYGSEEADTIRRLGHSGIGTPA
           orf5ng
50
                        RARRKSPYRRFAVHRRTRRQPPPAYADGDPREVSX----RRFCTV 131
           orf5
```

The complete strain B and gonococcal sequences (ORF5-1 & ORF5ng-1) show 92.4% identity in

RARRKSPYRRFAVHRRPRRQPPPAHADGDPREVSRACPHRRFCTV

55 304 aa overlap:

orf5ng

10 20 30 40 50 60 orf5ng-1.pep MDGAQPKTNFFERLIARLAREPDSAEDVINLLRQAHEQEVFDADTLTRLEKVLDFAELEV

		1111111111111111	1111111111			11 1111111	1::111	
	orf5-1	MDGAQPKTNFFERLI 10	ARLAREPDSI 20	AEDVLNLLRQA 30	AHEQEVFDAD 40	TLLRLEKVLI 50	60	
5	orf5ng-1.pep	70 RDAMITRSRMNVLKE	80 NDSIERITA	90 YVIDTAHSRFI	100 PVIGEDKDEV	110 LGILHAKDLI	120 LKYMFNP	
	- •		1111111111	111111111	[	111111111		
	orf5-1	70	80	90	100	110	120	
10		130 EOFHLKSVLRPAVFV	140	150	160	170	180	
	orf5ng-1.pep	- 1111111:1111111		11111111111	111111111		[]]]]]	
15	orf5-1	EQFHLKSILRPAVEV	PEGKSLTAL 140	LKEFREQRNHI 150	MAIVIDEYGO 160	170	180	
		190	200	210	220	230	240	
	orf5ng-1.pep	DIEDEFDEDESADDI	11:11:111	111111111111111111111111111111111111111	1:111111:		111 : 11	
20	orf5-1	EIEDEFDEDDSADNI 190	THAVSSERWR 200	IHAATEIEDI 210	NTFFGTEYSS 220	SEEADTIRP-( 230	SHSRVGT	
		250	260	270	280	290	300	
25	orf5ng-1.pep	PARARRKSPYRRFA	VHRRPRRQPP	PAHADGDPRE	VSRACPTAV:	SAQFRMTVRS	FSVSIRP	
23	orf5-1	SARARRKSPYRRFA 240 250	VHRRTRROPP 260	PAYADGDPRE 270	VSTAV	SAQFRMTVRA	FSVSIRP 90	
	•	240 230	200			-		
30	orf5ng-1.pep	IRQTX 					•	
	orf5-1	IRQTX 300						į
	Computer analysis		eid sequenc	ces indicate	es a putati	ve leader s	sequence, a	ınd
25	•						. ,	
35	identified the follow	mg nomotogies.						
						_		
	Homology with hem							
	Homology with hem ORF5 and TlyC prot							
	ORF5 and TlyC protoners 2 HM	teins show 58% aa	identity in	77 aa over	lap (BLAS	Tp). DTWRIHAATE		
40	ORF5 and TlyC prot	teins show 58% aa	identity in	177 aa over EIEDEFDEDDS HIEDEFDE++	lap (BLAS SADNIHAVSS AD I +S	Tp).  DTWRIHAATE  T+ + A T+	I+D	
40	ORF5 and TlyC prof	teins show 58% aa AIVIDEYGGTSGLVTF AIV+DE+G SGLVT AIVVDEFGAVSGLVTI ITFFGTEYSIEEADTI	identity in EDIIEQIVGE EDI+EQIVGE EDILEQIVGE	177 aa over EIEDEFDEDDS HIEDEFDE++	lap (BLAS SADNIHAVSS AD I +S	Tp).  DTWRIHAATE  T+ + A T+	I+D	
40	ORF5 and TlyC protons ORF5 2 HM HM TlyC 166 HM ORF5 62 IN	teins show 58% aa AIVIDEYGGTSGLVTF AIV+DE+G SGLVT AIVVDEFGAVSGLVTI	identity in EDIIEQIVGE EDILEQIVGE EDILEQIVGE 78	177 aa over EIEDEFDEDDS HIEDEFDE++	lap (BLAS SADNIHAVSS AD I +S	Tp).  DTWRIHAATE  T+ + A T+	I+D	
40 45	ORF5 and TlyC prof	teins show 58% aa laivideyggtsglvtf laiv+de+g sglvt laivvdefgavsglvti ltffgteysieeadti l f t++ ee dti laqfntdfddeevdti	identity in EDITEQIVGE EDITEQIVGE EDITEQIVGE 78 241	177 aa over EIEDEFDEDDS -IEDEFDE++ DIEDEFDEEEI	lap (BLAS SADNIHAVSS AD I +S	Tp).  DTWRIHAATE  T+ + A T+	I+D	
	ORF5 and TlyC protocolors  ORF5 2 HM HM TlyC 166 HM ORF5 62 IN TlyC 225 FN ORF5ng-1 also show	teins show 58% aa  AIVIDEYGGTSGLVTF  AIV+DE+G SGLVT  AIVVDEFGAVSGLVTI  OF T++ EE DTI  AQFNTDFDDEEVDTI  WS significant hom  Init1: 301 Init	identity in EDITEQIVER EDITEQIVER EDITEQIVER 78 241 alongy with	177 aa over EIEDEFDEDDS HEDEFDE++ DIEDEFDEEEI TTlyC:	lap (BLAS SADNIHAVSS AD I +S CAD-IRQLSR	Tp). DTWRIHAATE T+ + A T+ HTYAVRALTE	I+D	
	ORF5 and TlyC protocolors  ORF5 2 HM HM TlyC 166 HM ORF5 62 IN TlyC 225 FN ORF5ng-1 also show	teins show 58% aa laivideyggtsglvtf laiv+de+g sglvt laivvdefgavsglvti ltffgteysieeadti l f t++ ee dti laofntdfddeevdti ws significant hom	identity in EDITEQIVER EDITEQIVER 78 241 cology with 11: 419 ( 45.9% ide	a 77 aa over EIEDEFDEDDS FIEDEFDE++ DIEDEFDEEEI TTIyC: Opt: 668 entity in 2	lap (BLAS SADNIHAVSS AD I +S CAD-IRQLSR	Tp).  DTWRIHAATE  T+ + A T+  HTYAVRALTE	I+D IDD 224	
	ORF5 and TlyC protocolors  ORF5 2 HM HM TlyC 166 HM ORF5 62 IN TlyC 225 FN ORF5ng-1 also show	teins show 58% aa  AIVIDEYGGTSGLVTF  AIV+DE+G SGLVT  AIVVDEFGAVSGLVTI  ATFFGTEYSIEEADTI  A F T++ EE DTI  AQFNTDFDDEEVDTI  WS significant hom  Anit1: 301 Init  an score: 668;	ridentity in EDITEQIVGE EDITEQIVGE 78 241 cology with 45.9% ide 10 ENFFERLIAR	177 aa over EIEDEFDEDDS FIEDEFDE++ DIEDEFDEEEI TTlyC: Opt: 668 entity in 2	lap (BLAS) SADNIHAVSS AD I +S CAD-IRQLSR  242 aa ove 30 DVLNLLRQAF	Tp).  DTWRIHAATE  T+ + A T+  HTYAVRALTE  Tlap  40  EQEVFDADTE	I+D IDD 224 50 TRLEK	
45	ORF5 and TlyC protocolors  ORF5 2 HM HM TlyC 166 HM ORF5 62 IN TlyC 225 FN ORF5ng-1 also show SCORES IN Smith-Waterman	teins show 58% aa  IAIVIDEYGGTSGLVTF IAIV+DE+G SGLVT IAIVVDEFGAVSGLVTI ITFFGTEYSIEEADTI IAQFNTDFDDEEVDTI IVS significant hom Init1: 301 Init IAN SCORE: 668;  MDGAQPKT MNDEQQNSNQSENTKE	redileques EDILEQUES EDILEQUES 78 241 cology with 10 CNFFERLIAR	TlyC:  Opt: 668 entity in 20  LAR-EPDSAEI:	lap (BLAS SADNIHAVSS AD I +S SAD-IRQLSR 242 aa ove 30 DVLNLLRQAF ::::: :::	Tp).  DTWRIHAATE T+ + A T+ HTYAVRALTE  Prlap  40 HEQEVFDADTI ::::: :	I+D IDD 224 50 TRLEK :  REMIEG	
45	ORF5 and TlyC protocolors  ORF5 2 HM HM TlyC 166 HM ORF5 62 IN TlyC 225 FN ORF5ng-1 also show SCORES IN Smith-Waterman	teins show 58% aa  IAIVIDEYGGTSGLVTF IAIV+DE+G SGLVT IAIVVDEFGAVSGLVTI ITFFGTEYSIEEADTI IAQFNTDFDDEEVDTI INS significant hom Init1: 301 Init IAN SCORE: 668;  MDGAQPKT MDGAQPKT MNDEQQNSNQSENTKR 10	redileques  EDILEQIVE  EDILEQIVE  78  241  cology with  10  ENFFERLIAR     :  ::   EFFFQSLFGR  20	a 77 aa over ELEDEFDEDDS HEDEFDE++ DIEDEFDEEEI  TTlyC: Opt: 668 entity in 2  20 LAR-EPDSAEI :: :  FFQGELKNREI 30	lap (BLAS  GADNIHAVSS AD I +S GAD-IRQLSR  242 aa ove 30 DVLNLLRQAF ::::::::: ELVEVIRDSE 40	Tp).  DTWRIHAATE T+ + A T+ HTYAVRALTE  Prlap  40 HEQEVFDADTH :::::  :  LQNDLIDQNTH	T+D IDD 224  50 TRLEK :  REMIEG 60	
45	ORF5 and TlyC proton or Some state of Smith-Waterman or Sng-1.pep	teins show 58% aa  AIVIDEYGGTSGLVTF  IAIV+DE+G SGLVT  IAIVVDEFGAVSGLVTI  ITFFGTEYSIEEADTI  I F T++ EE DTI  IAQFNTDFDDEEVDTI  WS significant hom  Initl: 301 Init  An score: 668;  MDGAQPKT  MNDEQQNSNQSENTKE  10  60  VLDFAELEVRDAMITE	tidentity in EDIIEQIVGE EDIIEQIVGE 78 241 tology with 45.9% ide 10 FNFFERLIAR   : :  CPFFQSLFGR 20 70 RSRMNVLKEN	a 77 aa over EIEDEFDEDDS FIEDEFDE++ DIEDEFDEEEI  TTlyC: Opt: 668 entity in 2  20 LAR-EPDSAEI : :   :   FFQGELKNREI 30  80 DSIERITAYV	lap (BLAS SADNIHAVSS AD I +S SAD-IRQLSR  242 aa ove 30 SULLILIQAE SELVEVIRDSE 40 90 IDTAHSRFP	Tp).  DTWRIHAATE  T+ + A T+  HTYAVRALTE  A1  A2  A2  A3  A40  MEQEVFDADTI  MEQEVFDATI  MEDITAL  MEDI	J+D JDD 224 50 JTRLEK :  REMIEG 60 109 /LGILH	
45	ORF5 and TlyC proton or Some state of Smith-Waterman or Sng-1.pep	teins show 58% aa  AIVIDEYGGTSGLVTF IAIV+DE+G SGLVT IAIVVDEFGAVSGLVTI ITFFGTEYSIEEADTI IF T++ EE DTI IAQFNTDFDDEEVDTI WS significant hom  Anitl: 301 Init An score: 668;  MDGAQPKT MNDEQQNSNQSENTKE 10  60  VLDFAELEVRDAMITE  :::  :	identity in EDIIEQIVGE EDIIEQIVGE TREDILEQIVGE TREDILEQUE	TlyC:  Opt: 668 entity in 2  LAR-EPDSAE: :   :   FFQGELKNRE: 30  80 DSIERITAYV :::: : QDLNTCLNTI	lap (BLAS  BADNIHAVSS  AD I +S  BAD-IRQLSR  242 aa ove  30  DVLNLLRQAF  SELVEVIRDSE  40  90  IDTAHSRFPN   ::        IESAHSRFPN	Tp).  DTWRIHAATE  T+ + A T+  HTYAVRALTE  A1  A2  A2  A3  A40  MEQEVFDADTI  SCONDLIDQNTE  50  100  VIGEDKDEV   ::: : :   VIADADDRDN	J+D JDD 224  50 JTRLEK SEMIEG 60 109 VLGILH SISTERIS	
45	ORF5 and TlyC proton or Scores I smith-Waterman orf5ng-1.pep tlyc_haein	teins show 58% aa  IAIVIDEYGGTSGLVTF IAIV+DE+G SGLVT IAIVVDEFGAVSGLVTI ITFFGTEYSIEEADTI IAQFNTDFDDEEVDTI IAQFNTDFDDEEVDTI IVS significant hom Init1: 301 Init IAN SCORE: 668;  MDGAQPKT INDEQQNSNQSENTKE 10  60  VLDFAELEVRDAMITE  :::  :         VMEIAELRVRDIMIPE 70	redileques EDILEQUES EDILEQUES T8 241 cology with tn: 419 0 45.9% ide 10 CONFFERLIAR 11: 1::1 COPFFQSLFGR 20 70 RSRMNVLKEN 11::::: RSQLIFIEDQ 80	a 77 aa over ELEDEFDEDDS HIEDEFDE++ DIEDEFDEEEI  TTlyC: Opt: 668 entity in 2  LAR-EPDSAEI : :   :   FFQGELKNREI 30  80 DSIERITAYV :::: : : QOLNTCLNTI 90	lap (BLAS SADNIHAVSS AD I +S LAD-IRQLSR  242 aa ove 30 DVLNLLRQAF :::::::: 40 90 IDTAHSRFPN  ::	Tp).  DTWRIHAATE T+ + A T+ HTYAVRALTE  Prlap  40 HEQEVFDADTE SO  100 VIGEDKDEV  :::  : : //IADADDRDN: 110	J+D JDD 224 50 JTRLEK :  REMIEG 60 109 //LGILH	
45 50 55	ORF5 and TlyC proton or Scores or Smith-Waterman or Sng-1.pep tlyc_haein	teins show 58% aa  IAIVIDEYGGTSGLVTF IAIV+DE+G SGLVT IAIVVDEFGAVSGLVTI ITFFGTEYSIEEADTI ITFFGTEYSIEEADTI IAQFNTDFDDEEVDTI INS significant hom Init1: 301 Init IAN SCORE: 668;  MDGAQPKT INMODEQONSNQSENTKE 10 60 VLDFAELEVRDAMITE  :::  :         VMEIAELRVRDIMIPE 70 10 120 AKDLLKYMF-NPEOFI	tidentity in  EDIIEQIVGE EDI+EQIVGE EDILEQIVGE 78 241 tology with 210 10 ENFFERLIAR 11: 1::1 EXPFFQSLFGR 20 70 RSRMNVLKEN 11: :::: RSQIIFIEDQ 80 130 HLKSVLRPAV	a 77 aa over ELEDEFDEDDS HEDEFDE++ DIEDEFDEEEI  TTlyC:  Opt: 668 entity in 2  20 LAR-EPDSAEI ::   :   FFQGELKNREI 30  80 DSIERITAYV :::: : ODLNTCLNTI 90  140 FFVPEGKSLTA	lap (BLAS  AD I +S  A	Tp).  DTWRIHAATE T+ + A T+ HTYAVRALTE  Prlap  40 HEQEVFDADTI EQUIDIDQNTE 50  100 VIGEDKDEV HI::  : : VIADADDRDNI 110  160 NHMAIVIDEY	J+D JDD 224  50 JTRLEK :  REMIEG 60  109 VLGILH ::	
45 50 55	ORF5 and TlyC proton or Scores or Smith-Waterman or Sng-1.pep tlyc_haein	teins show 58% aa  IAIVIDEYGGTSGLVTF IAIV+DE+G SGLVT IAIVVDEFGAVSGLVTI ITFFGTEYSIEEADTI ITFFGTEYSIEEADTI IAQFNTDFDDEEVDTI INS significant hom Init1: 301 Init IAN SCORE: 668;  MDGAQPKT MNDEQONSNQSENTKE 10  60 VLDFAELEVRDAMITE  :::   :       VMEIAELRVRDIMIPE 70  10 120	identity in EDIIEQIVGE EDIIEQIVGE EDILEQIVGE 78 241 sology with En: 419 ( 45.9% ide 10 ENFFERLIAR   ::::: EFFFQSLFGR 20 70 RSRMNVLKEN   ::::: RSQIIFIEDQ 80 130 HLKSVLRPAV	a 77 aa over EIEDEFDEDDS HEDEFDE++ DIEDEFDEEEI  TTlyC: Opt: 668 entity in 2  20 LAR-EPDSAEI :: :  FFQGELKNREI 30  80 DSIERITAYV :::::::: 90  140 FFVPEGKSLTA :  : :	lap (BLAS  ADNIHAVSS AD I +S	Tp).  DTWRIHAATE T+ + A T+ HTYAVRALTE  Prlap  40 HEQEVFDADTI :::::  :  LQNDLIDQNTH 50  100 VIGEDKDEV  ::  : : VIADADDRDN: 110  160 NHMAIVIDEYV     :  :	J+D JDD 224  50 JTRLEK :  REMIEG 60  109 //LGILH ::     IVGILH 120  GGTSGL	

			130	140	150	160	170	180
5	orf5ng-1.pep	:   : VTIEDIL	111111111	[[]]:[	200 IHSVSAERWR  :::  : : IRQLSRHTYAV 210	::   : : :	: :: :	H‡L
10	orf5ng-1.per	11 1	: :1 .1 1	:	260 AVHRRPRRQP FKVTSADSRR 270			

15 Homology with a hypothetical secreted protein from *E.coli*:

ORF5a shows homology to a hypothetical secreted protein from E.coli:

```
SP|P77392|YBEX_ECOLI HYPOTHETICAL 33.3 KD PROTEIN IN CUTE-ASNB INTERGENIC REGION
          >qi|1778577 (UB2598) similar to H. influenzae [Escherichia coli] >qi|1786879
          (AE000170) f292; This 292 aa ORF is 23% identical (9 gaps) to 272 residues of an
          approx. 440 aa protein YTFL_HAEIN SW: P44717 [Escherichia coli] Length = 292
20
           Score = 212 bits (533), Expect = 3e-54
           Identities = 112/230 (48%), Positives = 149/230 (64%), Gaps = 3/230 (1%)
                      DGAQPKTNFXXRLIARLAR-EPDSAEDVLTLLRQAHEQEVFDADTLLRLEKVLDFSDLEV 60
          Query: 2
25
                                 L+++L EP + +++L L+R + + ++ D DT
          Sbjct: 10 DTISNKKGFFSLLLSQLFHGEPKNRDELLALIRDSGQNDLIDEDTRDMLEGVMDIADQRV 69
          Query: 61 RDAMITRSRMNVLKENDSIERITAYVIDTAHSRFPVIGEDKDEVLGILHAKDLLKYM-FN 119
                      RD MI RS+M LK N +++
                                              +I++AHSRFPVI EDKD + GIL AKDLL +M
30
           Sbjct: 70 RDIMIPRSQMITLKRNQTLDECLDVIIESAHSRFPVISEDKDHIEGILMAKDLLPFMRSD 129
           Query: 120 PEQFHLKSILRPAVFVPEGKSLTALLKEFREQRNHMAIVIDEYGGTSGLVTFEDIIEQIV 179
                       E F + +LR AV VPE K + +LKEFR QR HMAIVIDE+GG SGLVT EDI+E IV
           Sbjct: 130 AEAFSMDKVLRQAVVVPESKRVDRMLKEFRSQRYHMAIVIDEFGGVSGLVTIEDILELIV 189
35
           Query: 180 GDIEDEFDEDESADNIHAVSAERWRIHAATEIEDINAFFGTEYSSEEADT 229
                      G+IEDE+DE++ D
                                       +S
                                            W + A
                                                    IED N FGT +S EE DT
           Sbjct: 190 GEIEDEYDEEDDID-FRQLSRHTWTVRALASIEDFNEAFGTHFSDEEVDT 238
```

Based on this analysis, including the amino acid homology to the TlyC hemolysin-homologue from H. influenzae (hemolysins are secreted proteins), it was predicted that the proteins from N.meningitidis and N.gonorrhoeae are secreted and could thus be useful antigens for vaccines or diagnostics.

ORF5-1 (30.7kDa) was cloned in the pGex vector and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 2A shows the results of affinity purification of the GST-fusion protein. Purified GST-fusion protein was used to immunise mice, whose sera were used for Western blot analysis (Figure 1B). These experiments confirm that ORF5-1 is a surface-exposed protein, and that it is a useful immunogen.

#### Example 5

45

- The following partial DNA sequence was identified in N. meningitidis <SEQ ID 29>:
  - 1 ATGCGCGGCG GCAGGCCGGA TTCCGTTACC GTGCAGATTA TCGAAGGTTC
  - 51 GCGTTTTTCG CATATGAGGA AAGTCATCGA CGCAACGCCC GACATCGGAC

yceg

```
ACGACACCAA AGGCTGGAGC AATGAAAAAC TGATGGCGGA AGTTGCGCCC
                      GATGCCTTCA GCGGCAATCC TGAAGGGCAG TTTTTCCCCG ACAGCTACGA
               101
               151
                     AATCGATGCG GGCGGCAGTG ATTTGCAGAT TTACCAAACC GCCTACAAGG
                      GCGATGCAAC GCCGCCTGAA TGAQGGCATG GGAAAGCAGG CAGGACGGGC
                      TGCCTTATAA AAACCCTTAT GAAATGCTGA TTATGGCGAr CCTGGTCGAA
               251
                      AAGGAAACAG GGCATGAAGC CGASCSCGAC CATGTcGCTT CCGTCTTCGT
                301
5
                351
                      CAACCGCCTG AAAATCGGTA TGCGCCTGCA AACCGASSCG TCCGTGATTT
                401
                      ACGGCATGGG TGCGGCATAC AAGGGCAAAA TCCGTAAAGC CGACCTGCGC
                451
                      CGCGACACGC CGTACAACAC CTACACGCGC GGCGGTCTGC CGCCAACCCC
                501
                      GATTGCGCTG CCC..
                551
10
     This corresponds to the amino acid sequence <SEQ ID 30; ORF7>:
                      MRGGRPDSVT VQIIEGSRFS HMRKVIDATP DIGHDTKGWS NEKLMAEVAP
                      DAFSGNPEGQ FFPDSYEIDA GGSDLQIYQT AYKAMQRRLN EAWESRQDGL
                 51
                      PYKNPYEMLI MAXLVEKETG HEAXXDHVAS VFVNRLKIGM RLQTXXSVIY
                      GMGAAYKGKI RKADLRRDTP YNTYTRGGLP PTPIALP..
15
      Further sequence analysis revealed the complete DNA sequence <SEQ ID 31>:
                      ATGTTGAGAA AATTGTTGAA ATGGTCTGCC GTTTTTTTGA CCGTGTCGGC
                      AGCCGTTTTC GCCGCGCTGC TTTTTGTTCC TAAGGATAAC GGCAGGGCAT
                  51
                      ACCGAATCAA AATTGCCAAA AACCAGGGTA TTTCGTCGGT CGGCAGGAAA
                 101
                       CTTGCCGAAG ACCGCATCGT GTTCAGCAGG CATGTTTTGA CGGCGGCGGC
                 151
                       CTACGTTTTG GGTGTGCACA ACAGGCTGCA TACGGGGACG TACAGATTGC
 20
                 201
                       CTTCGGAAGT GTCTGCTTGG GATATCTTGC AGAAAATGCG CGGCGGCAGG
                 251
                       CCGGATTCCG TTACCGTGCA GATTATCGAA GGTTCGCGTT TTTCGCATAT
                 301
                       GAGGAAAGTC ATCGACGCAA CGCCCGACAT CGGACACGAC ACCAAAGGCT
                 351
                       GGAGCAATGA AAAACTGATG GCGGAAGTTG CGCCCGATGC CTTCAGCGGC
                 401
                       AATCCTGAAG GGCAGTTTTT CCCCGACAGC TACGAAATCG ATGCGGGCGG
 25
                 451
                       CAGTGATTTG CAGATTTACC AAACCGCCTA CAAGGCGATG CAACGCCGCC
                       TGAATGAGGC ATGGGAAAGC AGGCAGGACG GGCTGCCTTA TAAAAACCCCT
                 501
                       TATGAAATGC TGATTATGGC GAGCCTGGTC GAAAAGGAAA CAGGGCATGA
                  551
                  601
                       AGCCGACCGC GACCATGTCG CTTCCGTCTT CGTCAACCGC CTGAAAATCG
                  651
                       GTATGCGCCT GCAAACCGAC CCGTCCGTGA TTTACGGCAT GGGTGCGGCA
 30
                  701
                       TACAAGGGCA AAATCCGTAA AGCCGACCTG CGCCGCGACA CGCCGTACAA
                  751
                       CACCTACACG CGCGGCGGTC TGCCGCCAAC CCCGATTGCG CTGCCCGGCA
                  801
                       AGGCGGCACT CGATGCCGCC GCCCATCCGT CCGGCGAAAA ATACCTGTAT
TTCGTGTCCA AAATGGACGG CACGGGCTTG AGCCAGTTCA GCCATGATTT
                  851
                       GACCGAACAC AATGCCGCCG TCCGCAAATA TATTTTGAAA AAATAA
  35
                  901
       This corresponds to the amino acid sequence <SEQ ID 32; ORF7-1>:
                       MLRKLLKWSA VFLTVSAAVF AALLFVPKDN GRAYRIKIAK NQGISSVGRK
                       LAEDRIVFSR HVLTAAAYVL GVHNRLHTGT YRLPSEVSAW DILQKMRGGR
                        PDSVTVQIIE GSRFSHMRKV IDATPDIGHD TKGWSNEKLM AEVAPDAFSG
                  101
                        NPEGOFFPDS YEIDAGGSDL QIYQTAYKAM QRRLNEAWES RQDGLPYKNP
  40
                        YEMLIMASLV EKETGHEADR DHVASVFVNR LKIGMRLQTD PSVIYGMGAA
                        YKGKIRKADL RRDTPYNTYT RGGLPPTPIA LPGKAALDAA AHPSGEKYLY
                        FVSKMDGTGL SQFSHDLTEH NAAVRKYILK K*
        Computer analysis of this amino acid sequence gave the following results:
  45
        Homology with hypothetical protein encoded by yceg gene (accession P44270) of H.influenzae
        ORF7 and yceg proteins show 44% aa identity in 192 aa overlap:
                         MRGGRPDSVTVQIIEGSRFSHMRKVIDATPDIGHDTKGWSNEKLMA----EVAPDAFSG 55
              ORF7
                                                            K SNE++ A
                                    V+ IEG F RK ++ P +
                     102 LNSGKEVQFNVKWIEGKTFKDWRKDLENAPHLVQTLKDKSNEEIFALLDLPDIGQNLELK 161
   50
              yceg
                     56 NPEGQFFPDSYEIDAGGSDLQIYQTAYKAMQRRLNEAWESRQDGLPYKNPYEMLIMAXLV 115
              ORF7
                                           +DL++ + + + M++ LN+AW R + LP NPYEMLI+A +V
                          N EG +PD+Y
                      162 NVEGWLYPDTYNYTPKSTDLELLKRSAERMKKALNKAWNERDEDLPLANPYEMLILASIV 221
              yceg
                      116 EKETGHEAXXDHVASVFVNRLKIGMRLQTXXSVIYGMGAAYKGKIRKADLRRDTPYNTYT 175
   55
              ORF7
                                      VASVF+NRLK M+LQT +VIYGMG Y G IRK DL
                                                                               TPYNTY
                      222 EKETGIANERAKVASVFINRLKAKMKLQTDPTVIYGMGENYNGNIRKKDLETKTPYNTYV 281
```

```
ORF7 176 RGGLPFTPIALF 187
GLPPTPIA+F
yceg 282 IDGLPPTPIAMP 293
```

### The complete length YCEG protein has sequence:

5	. 1	MKKFLIAILL	LILILAGVAS	<b>FSYYKMTEFV</b>	KTPVNVQADE	LLTIERGTTS
	51	SKLATLFEQE	KLIADGKLLP	YLLKLKPELN	KIKAGTYSLE	NVKTVQDLLD.
	101	LLNSGKEVQF	NVKWIEGKTF	KDWRKDLENA	PHLVQTLKDK	SNEEIFALLD
	151	LPDIGONLEL	KNVEGWLYPD	TYNYTPKSTD	LELLKRSAER	MKKALNKAWN
	201	ERDEDLPLAN	PYEMLILASI	VEKETGIANE	RAKVASVFIN	RLKAKMKLQT
10	251	DPTVIYGMGE	NYNGNIRKKD	LETKTPYNTY	VIDGLPPTPI	AMPSESSLQA
	301	VANPEKTDFY	YFVADGSGGH	KFTRNLNEHN	KAVQEYLRWY	RSQKNAK

### Homology with a predicted ORF from N. meningitidis (strain A)

ORF7 shows 95.2% identity over a 187aa overlap with an ORF (ORF7a) from strain A of N.

### 15 meningitidis:

						10	20 30	) ·
	orf7.pep				MRGGRPDS	SVTVQIIEGSR	FSHMRKVIDATI	₽
	٠,				11111111			l
	orf7a	AAYVLGVHN	IRLHTGTYRLP	SEVSAWDIL	QKMRGGRPDS	SVTVQIIEGSR	FSHMRKVIDATI	₽
20		70	80	90	100	110	120	
•	• •							
		. 4	10 5	0	60	70	80 90	0
	orf7.pep	DIGHDTKGW	ISNEKLMAEVA	PDAFSGNPE	GOFFPDSYE	IDAGGSDLQIY	QTAYKAMQRRL1	N
		11 111111		$\mathbf{H}\mathbf{H}\mathbf{H}\mathbf{H}\mathbf{H}$	11111111	111111111:11	1 111111111	1
25	orf7a	DIEHDTKG	ISNEKLMAEVA	PDAFSGNPE	GQFFPDSYE	IDAGGSDLRIY	QIAYKAMQRRLI	N
	•	130	140	150	160	170	180	
	•	. 10					40 15	
11	orf7.pep	EAWESRODO	SLPYKNPYEML	IMAXLVEKE	TGHEAXXDH	Vasvfvnrlki	GMRLQTXXSVI	Y
30		111111111		111 1:111		11111111	111111 111	1
	orf7a	<del>-</del> ,					GMRLQTDPSVI	Y
		190	200	210	220	230	240	
	•		50 17	1	. 80			
25	67		CIRKADLRRDT	-				
35	orf7.pep	GMGAAING	TKKADEKKUI	PINITIRGG	PERTETALE			
•		CVCDDVKC	(TDK) DI DDD		11111111	~~~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~		
	orf7a		KIRKADLRRDT				SGEKYLYFVSKI	M
		250	260	270	280	290	300	
40	orf7a	DOTGI.SOF	SHDLTEHNAAV	PKYTI.KKY				
, +0	ULITA	310	320	330				
		210	520	230				

### The complete length ORF7a nucleotide sequence <SEQ ID 33> is:

	1	ATGTTGAGAA	AATTGTTGAA	ATGGTCTGCC	GTTTTTTTGA	CCGTATCGGC
•	51	AGCCGTTTTC	GCCGCGCTGC	TTTTCGTCCC	TAAAGACAAC	GGCAGGGCAT
45	101	ACAGGATTAA	AATTGCCAAA	AACCAGGGTA	TTTCGTCGGT	<b>CGGCAGGAAA</b>
•	151	CTTGCCGAAG	ACCGCATCGT	GTTCAGCAGG	CATGTTTTGA	CGGCGGCGGC
	201	CTACGTTTTG	GGTGTGCACA	ACAGGCTGCA	TACGGGGACG	TACAGACTGC
	251	CTTCGGAAGT	GTCTGCTTGG	GATATCTTGC	AGAAAATGCG	CGGCGGCAGG
	301	CCGGATTCCG	TTACCGTGCA	GATTATCGAA	GGTTCGCGTT	TTTCGCATAT
50	351	GAGGAAAGTC	ATCGACGCAA	CGCCCGACAT	CGAACACGAC	ACCAAAGGCT
	401	GGAGCAATGA	AAAACTGATG	GCGGAAGTTG	CCCCTGATGC	CTTCAGCGGC
	451	AATCCTGAAG	GGCAGTTTTT	CCCCGACAGC	TACGAAATCG	<b>ATGCGGGCGG</b>
	501	CAGCGATTTA	CGGATTTACC	AAATCGCCTA	CAAGGCGATG	CAACGCCGAC
	551	TGAATGAGGC	ATGGGAAAGC	AGGCAGGACG	GGCTGCCTTA	TAAAAACCCT
55	601	TATGAAATGC	TGATTATGGC	GAGCCTGATC	GAAAAGGAAA	CAGGGCATGA
•	651	AGCCGACCGC	GACCATGTCG	CTTCCGTCTT	CGTCAACCGC	<b>CTGAAAATCG</b>
	701	GTATGCGCCT	GCAAACCGAC	CCGTCCGTGA	TTTACGGCAT	<b>GGGTGCGGCA</b>
	751	TACAAGGGCA	AAATCCGTAA	AGCCGACCTG	CGCCGCGACA	CGCCGTACAA
	801	CACCTACACG	CGCGGCGGTC	TGCCGCCAAC	CCCGATCGCG	CTGCCCGGCA
60	851	AGGCGGCACT	CGATGCCGCC	GCCCATCCGT	CCGGTGAAAA	ATACCTGTAT
	901	TTCGTGTCCA	AAATGGACGG	TACGGGCTTG	AGCCAGTTCA	GCCATGATTT
	951	GACCGAACAC	AACGCCGCCG	TTCGCAAATA	TATTTTGAAA	AAATAA

This is predicted to encode a protein having amino acid sequence <SEQ ID 34>:

```
1 MLRKLLKWSA VFLTVSAAVF AALLFVPKDN GRAYRIKIAK NQGISSVGRK
51 LAEDRIVFSR HVLTAAAYVL GVHNRLHTGT YRLPSEVSAW DILQKMRGGR
101 PDSVTVQIIE GSRFSHMRKV IDATPDIEHD TKGWSNEKLM AEVAPDAFSG
151 NPEGQFFPDS YEIDAGGSDL RIYQIAYKAM QRRLNEAWES RQDGLPYKNP
201 YEMLIMASLI EKETGHEADR DHVASVFVNR LKIGMRLQTD PSVIYGMGAA
251 YKGKIRKADL RRDTPYNTYT RGGLPPTPIA LPGKAALDAA AHPSGEKYLY
301 FVSKMDGTGL SQFSHDLTEH NAAVRKYILK K*
```

A leader peptide is underlined.

# 10 ORF7a and ORF7-1 show 98.8% identity in 331 aa overlap:

			10	20	30 EVPKDNGRAYI	40	50 SSVGRKTAED	60 RTVFSR
	orf7a.pep	1111111	11111111111	111111	111111111			11111
15	orf7-1	MLRKLLK	WSAVFLTVSA 10	AVFAALL 20	FVPKDNGRAYI 30	RIKIAKNQGIS 40	SSVGRKLAED 50	60
•			70	80	90	100	110	120
	orf7a.pep	HVLTAAA	YVLGVHNRLH	1111111	SEVSAWDILQ	11111111	:	. [ ] [ ] [ ] [ ]
20	orf7-1	HVLTAAZ	YVLGVHNRLI 70	TGTYRLP 80	SEVSAWDILQ	KMRGGRPDSV 100	TVQIIEGSRE 110	TSHMRKV 120
			130	140	150	160	170	180 '
25	orf7a.pep	111111		11111111		111111111		
23	orf7-1	IDATPD:	IGHDTKGWSNI 130	EKLMAEVA 140	PDAFSGNPEG 150	QFFPDSYEID 160	AGGSDLQIYO 170	2TAYKAM 180
			190	200	210	220	230	240
30	orf7a.pep	111111	ETTTELLITE	1111111	LIMASLIEKET	111111111	11111111	11111
	orf7-1	QRRLNE	AWESRODGLP	YKNPYEMI 200	IMASLVEKET 210	GHEADRDHVA 220	SVFVNRLKI 230	GMRLQTD 240
35			250	260	270	280	290	300
33	orf7a.pep	111311	11111111111	1111111	PYNTYTRGGI	1111111111	11111111	111111
	orf7-1	PSVIYG	MGAAYKGKIR 250	KADLRRD' 260	rpyntytrggi 270	PPTPIALPGE 280	ZALDAAAHP 290	300
40			310	320	330			
	orf7a.pep	FVSKMD	GTGLSQFSHD	LTEHNAA	VRKYILKKX 			
45	orf7-1		GTGLSQFSHD 310	320	VRKYILKKX 330			

## Homology with a predicted ORF from N. gonorrhoeae

ORF7 shows 94.7% identity over a 187aa overlap with a predicted ORF (ORF7.ng) from N. gonorrhoeae:

	80		
50	orf7	MRGGRPDSVTVQIIEGSRFSHMRKVIDATPDIGHDTKGWSNEKLMAEVAPDAFSGNPEGQ	
	orf7ng	MRGGRPDSVTVQIIEGSRFSHMRKVIDATPDIGHDTKGWSNEKLMAEVAPDAFSGNPEGC	60
55	orf7	FFPDSYEIDAGGSDLQIYQTAYKAMQRRLNEAWESRQDGLPYKNPYEMLIMAXLVEKETG	120
	orf7ng	FFPDSYEIDAGGSDLQIYQTAYKAMQRRLNEAWAGRQDGLPYKNPYEMLIMASLIEKETO	120
	orf7	HEAXXDHVASVFVNRLKIGMRLQTXXSVIYGMGAAYKGKIRKADLRRDTPYNTYTRGGLE	
60	orf7ng	HEADRDHVASVFVNRLKIGMRLQTDPSVIYGMGAAYKGKIRKADLRRDTPYNTYTGGGLI	180
	orf7	PTPIALP 19	17

orf7ng

60

|| |||||
PTRIALPGKAAMDAAAHPSGEKYLYFVSKMDGTGLSQFSHDLTEHNAAVRKYILKK 236

An ORF7ng nucleotide sequence <SEQ ID 35> is predicted to encode a protein having amino acid sequence <SEQ ID 36>:

	sequence <seq i<="" th=""><th>D 36&gt;:</th><th></th><th></th><th></th><th></th><th>:</th></seq>	D 36>:					:
5	1 51	MRGGRPDSVT VQ DAFSGNPEGQ FF	IIEGSRFS HN	RKVIDATP D	IGHDTKGWS N	EKLMAEVAP AWAGRODGI.	
	101	PYKNPYEMLI MA	SLIEKETG HE	ADRDHVAS V	FVNRLKIGM R	LOTOPSVIY	
	151	GMGAAYKGKI RK	ADLREDTP YN	TYTGGGLP P	TRIALPGKA A	MDAAAHPSG	
		EKYLYFVSKM DG					
10	Further sequence	analysis revea	led a partial	DNA seque	nce of ORF	7ng <seq ii<="" th=""><th>37&gt;:</th></seq>	37>:
	1	taccgaatca	AGATTGCCAA	AAATCAGGGT	ATTTCGTCGG	TCGGCAGGAA	
	51	ACTTGCcgaA	GACCGCATCG	TGTTCAGCAG	GCATGTTTTG	ACAGCGGCGG	
	101	CCTACGTTTT	GGGTGTGCAC	AACAGGCTGC	ATACGGGGAC	gTACAGATTG	
15	151 201	CCCCCATTCC	CTTACCCTCC	ACAMMAMCCA	CAGAAAATGC	GCGGCGGCAG	
13	201 251	GCCGGATTCC TGAGGAAAGT	CATCCACCCA	AGATTATUGA	AGGTTCGCGT	TTTTCGCATA	
	301	TGGAGCAATG	ANNACCCA	CCCCCAACA	CCCCCCCA	CACCAAAGGC	
	351	CAATCCTGAA	CCCCACTTOAL	TTCCCCACAC	CONTRACTANAMA	COTTCAGCGG	
	401	GCAGCGATTT	GCDCAGIIII CCDCDTTTDC	CARACCCCCCT	DCDDCCCCDM	CATGUGGGCG	
20	451	CTGAACGAGG	CATCCCCACC	CARACCGCCI	CCCCTCCCTT	ATTARANCECCEC	
20	501	TTATGAAATG	CTCATTATCC	CAGGCAGGAC	CCAAAACCAA	ATAMAMACCC	
	551	AGGCCGACCG	CCACCATCTC	CCTTCCCTCT	TCCTCAACCC	CCCCAAAAA	
	601	GGTATGCGCC	TGCAAACCGA	CCCGTCCGTG	ATTTACCCCA	TECETECECE	
	651	ATACAAGGGC	AAAATCCGTA	AAGCCGACCT	GCGCCGCGAC	ACCCCCTACA	
25	701	aCAccTAtac	aaacaaaaac	ttgccgccaa	cccagattac	actacceaac	
	751	Aaggcggcaa	tagataccac	cacccaccca	tccggcgaAa	aatacctcTa	
	801	tttcgtgtcC .	AAAATGGACG	GCACGGGCTT	GAGCCAGTTC	AGCCATGATT	
	851	TGACCGAACA					
	This corresponds	to the amino a	cid sequenc	e <seo 3<="" id="" th=""><th>38: ORF7ng</th><th>-1&gt;:</th><th></th></seo>	38: ORF7ng	-1>:	
	-		•	•	, ,		
30	1	YRIKIAKNQG	ISSVGRKLAE	DRIVESRHVL	TAAAYVLGVH	NRI HTGTYRI.	
	51	PSEVSAWDIL	OKMRGGRPDS	VTVOILEGSR	FSHMRKVIDA	TPDIGHDTKG	
	101	WSNEKLMAEV .	APDAFSGNPE	GOFFPDSYEI	DAGGSDLOIY	OTAYKAMORR	
	151	LNEAWAGRQD	GLPYKNPYEM	LIMASLIEKE	TGHEADRDHV	ASVFVNRLKI	
	201	GMRLQTDPSV	IYGMGAAYKG	KIRKADLRRD	TPYNTYTGGG	LPPTRIALPG	
35	251	KAAMDAAAHP	SGEKYLYFVS	KMDGTGLSQF	SHDLTEHNAA	VRKYILKK*	
	ORF7ng-1 and O	RF7-1 show 98	8.0% identit	y in 298 aa c	overlap:		
		10	20	30	40	50	60
	orf7-1.pep					SVGRKLAEDRIV	FCRHVI.
40	orf7ng-1					SVGRKLAEDRIV	
	_				10	20	30
	•	70	80	90	100	110	120
	orf7-1.pep					/QIIEGSRFSHM	120
45	011: 1:Pop	1111111111	1111111111	11111111111	HILLIIIII		IKKATDA
	orf7ng-1	TAAAYVI.GVH	NRI HTGTYRI.	SEVSAWDTLO	(MDCCDDDCtmm	/QIIEGSRFSHM	
		40		60	70	80	90
					. •		50
		130	140	150	160	170	180
50	orf7-1.pep	TPDIGHDTKG	wsneklmaev <i>i</i>	PDAFSGNPEG	OFFPDSYEIDAG	GSDLQIYQTAY	KAMORR
		[]]]]]	1				111111
	orf7ng-1	TPDIGHDTKG	wsneklmaev <i>i</i>		QFFPDSYEIDAG	GSDLQIYQTAY	KAMORR
	•	100	110	120	130	140	150
<i>E E</i>							
55		190	200	210	220	230	240
	orf7-1.pep		GLPYKNPYEMI	IMASLVEKETO	HEADRDHVAS	VFVNRLKIGMRI	QTDPSV
		11111 : 111	1	11111:1111			
	orf7ng-1					VFVNRLKIGMRI	
60		160	170	180	190	200	210
vv							

260

270

 ${\tt IYGMGAAYKGKIRKADLRRDTPYNTYTRGGLPPTPIALPGKAALDAAAHPSGEKYLYFVS}$ 

280

290

```
IYGMGAAYKGKIRKADLRRDTPYNTYTGGGLPPTRIALPGKAAMDAAAHPSGEKYLYFVS
          orf7ng-1
                                                  240
                                        230
                           310
                                     320
                                               330
5
                       KMDGTGLSQFSHDLTEHNAAVRKYILKKX
          orf7-1.pep
                        1111111111111111111111111111111111111
                       KMDGTGLSQFSHDLTEHNAAVRKYILKKX
          orf7ng-1
                                        290
                              280
10
     In addition, ORF7ng-1 shows significant homology with a hypothetical E.coli protein:
          sp|P28306|YCEG_ECOLI HYPOTHETICAL 38.2 KD PROTEIN IN PABC-HOLB INTERGENIC REGION
          gi|1787339 (AE000210) 0340; 100% identical to fragment YCEG_ECOLI SW: P28306 but
          has 97 additional C-terminal residues [Escherichia coli] Length = 340
            Score = 79 (36.2 bits), Expect = 5.0e-57, Sum P(2) = 5.0e-57
15
            Identities = 20/87 (22%), Positives = 40/87 (45%)
                     10 GISSVGRKLAEDRIVFSRHVLTAAAYVLGVHNRLHTGTYRLPSEVSAWDILQKMRGGRPD 69
           Ouery:
                                                                  +++ ++L+ + G+
                                                           GTYR
                                         v
                        G ++G +L D+I+
                     49 GRLALGEQLYADKIINRPRVFQWLLRIEPDLSHFKAGTYRFTPQMTVREMLKLLESGKEA 108
           Sbjct:
20
                     70 SVTVOIIEGSRFSHMRKVIDATPDIGH 96
           Query:
                                       K +
                                             PIH
                           ++++EG R S
                    109 QFPLRLVEGMRLSDYLKQLREAPYIKH 135
           Sbjct:
 25
            Score = 438 (200.7 bits), Expect = 5.0e-57, Sum P(2) = 5.0e-57
            Identities = 84/155 (54%), Positives = 111/155 (71%)
                    120 EGQFFPDSYEIDAGGSDLQIYQTAYKAMQRRLNEAWAGRQDGLPYKNPYEMLIMASLIEK 179
           Query:
                        EG F+PD++ A +D+ + + A+K M + ++ AW GR DGLPYK+ +++ MAS+IEK
 30
                    158 EGWFWPDTWMYTANTTDVALLKRAHKKMVKAVDSAWEGRADGLPYKDKNQLVTMASIIEK 217
           Sbjct:
                    180 ETGHEADRDHVASVFVNRLKIGMRLQTDPSVIYGMGAAYKGKIRKADLRRDTPYNTYTGG 239
           Ouery:
                             ++RD VASVF+NRL+IGMRLQTDP+VIYGMG Y GK+ +ADL
                                                                          T YNTYT
                     218 ETAVASERDKVASVFINRLRIGMRLQTDPTVIYGMGERYNGKLSRADLETPTAYNTYTIT 277
 35
            Sbjct:
                     240 GLPPTRIALPGKAAMDAAAHPSGEKYLYFVSKMDG 274
            Query:
                         GLPP IA PG ++ AAAHP+
                                                YLYFV+
                     278 GLPPGATATPGADSLKAAAHPAKTPYLYFVADGKG 312
            Sbjct:
 40
```

Based on this analysis, including the fact that the *H.influenzae* YCEG protein possesses a possible leader sequence, it is predicted that the proteins from *N meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 6

The following partial DNA sequence was identified in N.meningitidis <SEQ ID 39>:

```
CGTTTCAAAA TGTTAACTGT GTTGACGGCA ACCTTGATTG CCGGACAGGT
                    ATCTGCCGCC GGAGGCGGTG CGGGGGATAT GAAACAGCCG AAGGAAGTCG
                51
                    GAAAGGTTTT CAGAAAGCAG CAGCGTTACA GCGAGGAAGA AATCAAAAAC
                    GAACGCGCAC GGCTTGCGGC AGTGGGCGAG CGGGTTAATC AGATATTTAC
                151
                    GTTGCTGGGA GGGGAAACCG CCTTGCAAAA GGGGCAGGCG GGAACGGCTC
50
                201
                    TGGCAACCTA TATGCTGATG TTGGAACGCA CAAAATCCCC CGAAGTCGCC
                251
                    GAACGCGCCT TGGAAATGGC CGTGTCGCTG AACGCGTTTG AACAGGCGGA
                301
                    AATGATTTAT CAGAAATGGC GGCAGATTGA GCCTATACCG GGTAAGGCGC
                351
                    AAAAACGGGC GGGGTGGCTG CGGAACGTGC TGAGGGAAAG AGGAAATCAG
                401
                    CATCTGGACG GACGGGAAGA AGTGCTGGCT CAGGCGGACG AAGGACAG
55
```

This corresponds to the amino acid sequence <SEQ ID 40; ORF9>:

1 ..RFKMLTVLTA TLIAGQVSAA GGGAGDMKQP KEVGKVFRKQ QRYSEEEIKN 51 ERARLAAVGE RVNQIFTLLG GETALQKGQA GTALATYMLM LERTKSPEVA 101 ERALEMAVSL NAFEQAEMIY QKWRQIEPIP GKAQKRAGWL RNVLRERGNQ

#### 151 HLDGREEVLA QADEGQ

#### Further sequence analysis revealed the complete DNA sequence <SEQ ID 41>:

```
ATGTTACCTA ACCGTTTCAA AATGTTAACT GTGTTGACGG CAACCTTGAT
                51
                    TGCCGGACAG GTATCTGCCG CCGGAGGCGG TGCGGGGGAT ATGAAACAGC
5
               101
                    CGAAGGAAGT CGGAAAGGTT TTCAGAAAGC AGCAGCGTTA CAGCGAGGAA
                    GAAATCAAAA ACGAACGCGC ACGGCTTGCG GCAGTGGGCG AGCGGGTTAA
               151
                    TCAGATATTT ACGTTGCTGG GAGGGGAAAC CGCCTTGCAA AAGGGGCAGG
               201
               251
                    CGGGAACGC TCTGGCAACC TATATGCTGA TGTTGGAACG CACAAAATCC
                    CCCGAAGTCG CCGAACGCGC CTTGGAAATG GCCGTGTCGC TGAACGCGTT
               301
10
                    TGAACAGGCG GAAATGATTT ATCAGAAATG GCGGCAGATT GAGCCTATAC
               351
                    CGGGTAAGGC GCAAAAACGG GCGGGGTGGC TGCGGAACGT GCTGAGGGAA
               401
                    AGAGGAAATC AGCATCTGGA CGGACTGGAA GAAGTGCTGG CTCAGGCGGA
               451
                    CGAAGGACAG AACCGCAGGG TGTTTTTATT GTTGGCACAA GCCGCCGTGC
               501
                    AACAGGACGG GTTGGCGCAA AAAGCATCGA AAGCGGTTCG CCGCGCGGCG
               551
15
                601
                    TTGAAATATG AACATCTGCC CGAAGCGGCG GTTGCCGATG TGGTGTTCAG
                    CGTACAGGGA CGCGAAAAGG AAAAGGCAAT CGGAGCTTTG CAGCGTTTGG
                651
               701
                    CGAAGCTCGA TACGGAAATA TTGCCCCCCA CTTTAATGAC GTTGCGTCTG
                    ACTGCACGCA AATATCCCGA AATACTCGAC GGCTTTTTCG AGCAGACAGA
                751
                801
                    CACCCAAAAC CTTTCGGCCG TCTGGCAGGA AATGGAAATT ATGAATCTGG
20
                    TTTCCCTGCA CAGGCTGGAT GATGCCTATG CGCGTTTGAA CGTGCTGTTG
                851
                    GAACGCAATC CGAATGCAGA CCTGTATATT CAGGCAGCGA TATTGGCGGC
                901
               951
                    AAACCGAAAA GAAGGTGCTT CCGTTATCGA CGGCTACGCC GAAAAGGCAT
                    ACGGCAGGGG GACGGAGGAA CAGCGGAGCA GGGCGGCGCT AACGGCGGCG
              1001
                    ATGATGTATG CCGACCGCAG GGATTACGCC AAAGTCAGGC AGTGGCTGAA
              1051
25
              1101
                     AAAAGTATCC GCGCCGGAAT ACCTGTTCGA CAAAGGTGTG CTGGCGGCTG
                     CGGCGGCTGT CGAGTTGGAC GGCGGCAGGG CGGCTTTGCG GCAGATCGGC
               1151
              1201
                     AGGGTGCGGA AACTTCCCGA ACAGCAGGGG CGGTATTTTA CGGCAGACAA
                     TTTGTCCAAA ATACAGATGC TCGCCCTGTC GAAGCTGCCC GATAAACGGG
               1251
               1301
                     AGGCTTTGAG GGGGTTGGAC AAGATTATCG AAAAACCGCC TGCCGGCAGT
30
                     AATACAGAGT TACAGGCAGA GGCATTGGTA CAGCGGTCAG TTGTTTACGA
               1351
               1401
                     TCGGCTTGGC AAGCGGAAAA AAATGATTTC AGATCTTGAA AGGGCGTTCA
                     GGCTTGCACC CGATAACGCT CAGATTATGA ATAATCTGGG CTACAGCCTG
               1451
                     CTGACCGATT CCAAACGTTT GGACGAAGGT TTCGCCCTGC TTCAGACGGC
               1501
                     ATACCAAATC AACCCGGACG ATACCGCTGT CAACGACAGC ATAGGCTGGG
               1551
35
                     CGTATTACCT GAAAGGCGAC GCGGAAAGCG CGCTGCCGTA TCTGCGGTAT
               1601
               1651
                     TCGTTTGAAA ACGACCCCGA GCCCGAAGTT GCCGCCCATT TGGGCGAAGT
                     GTTGTGGGCA TTGGGCGAAC GCGATCAGGC GGTTGACGTA TGGACGCAGG
               1701
                     CGGCACACCT TACGGGAGAC AAGAAAATAT GGCGGGAAAC GCTCAAACGT
               1751
                     CACGGCATCG CATTGCCCCA ACCTTCCCGA AAACCTCGGA AATAA
               1801
      This corresponds to the amino acid sequence <SEQ ID 42; ORF9-1>:
40
```

	1	MLPNRFKMLT	VLTATLIAGQ	VSAAGGGAGD	MKQPKEVGKV	FRKQQRYSEE
	51	EIKNERARLA	AVGERVNQIF	TLLGGETALQ	KGQAGTALAT	YMLMLERTKS
	101	PEVAERALEM	AVSLNAFEQA	EMIYQKWRQI	EPIPGKAQKR	AGWLRNVLRE
	151	RGNQHLDGLE	EVLAQADEGQ	NRRVFLLLAQ	AAVQQDGLAQ	KASKAVRRAA
45	201	LKYEHLPEAA	VADVVFSVQG	REKEKAIGAL	QRLAKLDTEI	LPPTLMTLRL
	251	TARKYPEILD	GFFEQTDTQN	LSAVWQEMEI	MNLVSLHRLD	DAYARLNVLL
	301	ERNPNADLYI	QAAILAANRK	EGASVIDGYA	EKAYGRGTEE	QRSRAALTAA
	351	MMYADRRDYA	KVRQWLKKVS	APEYLFDKGV	LAAAAAVELD	GGRAALRQIG
	401	RVRKLPEQQG	RYFTADNLSK	IQMLALSKLP	DKREALRGLD	KIIEKPPAGS
50	451	NTELQAEALV	QRSVVYDRLG	KRKKMISDLE	RAFRLAPDNA	QIMNNLGYSL
	501	LTDSKRLDEG	FALLQTAYQI	NPDDTAVNDS	IGWAYYLKGD	AESALPYLRY
	551	SFENDPEPEV	AAHLGEVLWA	LGERDQAVDV	WTQAAHLTGD	KKIWRETLKR
	601	HGIALPOPSR	KPRK*			

Computer analysis of this amino acid sequence gave the following results:

#### 55 Homology with a predicted ORF from *N. meningitidis* (strain A)

ORF9 shows 89.8% identity over a 166aa overlap with an ORF (ORF9a) from strain A of N. meningitidis:

		10	20	30	40	50
	orf9.pep	RFKMLTVLTATLIAG	QVSAAGGG	agdmkqpkevgi	KVFRKQQRYS	EEEIKNERARLA
60	•	11 :1:11:1:1:1:1	1: 11 11	1:1   11111	[]]	111111111111
	orf9a	MLPARFTILSVLAAALLAG	Qayaagi	aadakppkevgi	KVFRKQQRYS	EEEIKNERARLA

		1	0 20	3	10 40	50
		60	70	80	90 1	.00 110
_	orf9.pep		14 1 1 1 1 1 1 1 1			VAERALEMAVSLNAFEQA
5	orf9a	AVGERVNQI 60	FTLLGXETALC	(KGQAGTALAT)	MLMLERTKSPE	AVEKAPENAASPINAEEAA
			130	140	150	160
10	orf9.pep	120 EMIYQKWRQ	TEPTPGKAOKI	RAGWLRNVLRE	RGNOHLDGREEV	/LAQADEGQ
	orf9a	 EMIYQKWRQ	<b>IEPIPGKAQK</b> I	RAGWLRNVLRE		(TAOADEXONKKA F PPPAO
		120		-	50 160	
15	orf9a	AAVQQDGLF 180	190 190	ALRYEHLPEAA 200 2	VADVVFSVQXRI 10 22	EKEKAIGALQRLAKLDTEI 0 230
	The complete len	igth ORF9a n	ucleotide se	quence <se< th=""><th>Q ID 43&gt; is:</th><th></th></se<>	Q ID 43> is:	
	1	ATGTTACCCG (	CCCGTTTCAC	CATTTTATCT	GTGCTCGCGG	CAGCCCTGCT
	51	TGCCGGGCAG (	CCTATCCCG	CCGGCGCGGC	GGATGCGAAG	CCGCCGAAGG
20	101	ANANACCANC (	CCCCACGGCT	TGCGGCAGTG	GGCGAGCGGG	TTAATCAGAT
	151 201	ADDUA COURC (	CTCCCANGGG	AAACCGCCTT	GCAAAAGGGG	CAGGCGGGAA
	251	CCCCMCMCCCC '	እጥሬጥልጥልጥር	CTGATGTTGG	AACGCACAAA	ATCCCCCGAA
	301	GTCGCCGAAC GGCGGAAATG	GCGCCTTGGA	AATGGCCGTG	GATTGAGCCT	ATACCGGGTA
25	351 401	ACCCCCAAAA	ACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	TGGCTGCGGA	ACGTGCTGAG	GGAAAGAGGA
	451	እ አመር አር ር እጥር	TACACCCACT	GGAAGAANTG	CTGGCTCAGG	CGGACGAANG
	501	ACAGAACCGC ACGGGTTGGC	AGGGTGTTTT	TATTGTTGGC	TTCGCCGCGC	GGCGTTGAGA
20	551 601	のみのぐみみぐみがぐ	TCCCCGAAGC	GGCGGTTGCC	GATGTGGTGT	TCAGCGTACA
30	651	CONTRACCOCRA	AACCAAAACG	CAATCGGAGC	TTTGCAGCGT	TTGGCGAAGC
	701	TCGATACGGA CGCAAATATC	AATATTGCCC	CCCACTTTAA	TGACGTTGCG	CAGACACCCA
	751 801	አ አ አ ሮርጥጥጥሮር	CCCCTCTGGC	AGGAAATGGA	AATTATGAAT	CTGGTTTCCC
35	851	mccacacacce	CCATCATCCC	TATGCGCGTT	TGAACGTGCT	GTTGGAACGC
55	901	እ አመርርር እ አጥር	CACACCTGTA	TATTCAGGCA	GCGATATTGG	CGGCAAACCG
	951	AAAAGAANGT	GCTTCCGTTA	CCACGCCTA	CGCCGAAAAG CAATGACGGC	GGCGATGATA
	1001 1051	mamccccacc	CAACCCATTA	CACCAAAGTC	AGGCAGTGGT	TGAAAAAGT
40	1101	CHCCCCCCCC	CNATACCTGT	TCGACAAAGG	TGTGCTGGCG	GCTGCGGCGG
	1151	CTGTCGAGTT	GGACNGCGGC	AGGGCGGCTT	TGCGGCAGAT TTTACGGCAG	ACAATTTGTC
	1201 1251	~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~	<b>አ</b> ሞርሞሞርርርርር	TGTCGAAGCT	GCCCGACAAA	CGGGAGGCTT
	1301	man coccert	CCACAACATT	ATCGAAAAAC	CGCCTGCCGG	CAGTAATACA
45	1351	CROMMACACC	ヘカヘカにこの内で	CCTACAGCGG	TCAGTTGTTT	ACGATCGGCT
	1401	TGGCAAGCGG	AAAAAAATGA	ATGAATAATC	TGGGCTACAG	TTCAGGCTTG CCTGCTTTCC
	1451 1501	ር አመጥርር አ አ አር	CTTTCGACGA	AGGCTTCGCC	CTGCTTCAGA	CGGCATACCA
	1551	A ARCA ACCCC	CACCATACCG	CTGTCAACGA	CAGCATAGGC	TGGGCGTATT
50	1601		CGACGCGGAA	AGCGCGCTGC	CGTATCTGCG	GTATTCGTTT AAGTGTTGTG
	1651 1701	CCCATTCCCC	CAACGCGATC	: AGGCGGTTGA	CGTATGGACG	CAGGCGGCAC
	1751	አርርጥተልርርርG	AGACAAGAAA	ATATGGCGGG	; AAACGCTCAA	ACGTCACGGC
	1801			CCGAAAACCT		
55	This encodes a					
	1	MLPARFTILS	VLAAALLAG	AYAAGAADAI	PPKEVGKVFF	R KQQRYSEEEI
	51		GERVNQIFTI	LGXETALQKO	O TOCKDOKRACE OF THE CONTRACT	IMLERTKSPE WLRNVLRERG
	101 151	MOUT BOT PEY	LYCAUETONI	RVFLLLAOA	A VOODGLAQKA	A SKAVRRAALR
60	201	VEUT DEA AUZ	DWFSVOXRI	. KEKAIGALOI	R LAKLDTEILI	PTLMTLRLTA
00	251	DEVERTIBLE	FEOTOTONI.	R AVWOEMEIM	N LVSLHRLDDA	A YARLNVLLER R GRAAMTAAMI
	301	マネロロロロマザビス	I POWIKKVSA	P EYLFDKGVL	A AAAAVELDX	3 RAALKQIGKV
	351 401	DAL DEVOCES	Z ETADNI.SKT	O MFALSKLPD	K REALRGLDK.	I IEKPPAGSNT
65	4 5 1	ET OVEVI MOI	STAYDRIGK	R KKMISDLER	A FRLAPDNAQ	T WNNTGISTTS
	501	DOKOT DECE	Α Τ.Τ.ΟΥΑΥΟΙΝ	P DDTAVNDSI	<b>G WAYYLKXDA</b>	E SALPYLRYSF K IWRETLKRHG
	551	ENDPEPEVA	A HTGEATMWT	G EYNĀWANAM	T AUTHOUS	

#### 601 IALPQPSRKP RK\*

### ORF9a and ORF9-1 show 95.3% identity in 614 aa overlap:

5	orf9a.pep	10 20 30 40 50 MLPARFTILSVLAAALLAGQAYAAGAADAKPPKEVGKVFRKQQRYSEEEIKNERARLA
	orf9-1	
10	orf9a.pep	60 70 80 90 100 110 AVGERVNQIFTLLGXETALQKGQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFEQA
15	orf9-1	AVGERVNQIFTLLGGETALQKGQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFEQA 70 80 90 100 110 120
	orf9a.pep	120 130 140 150 160 170 EMIYQKWRQIEPIPGKAQKRAGWLRNVLRERGNQHLDGLEEXLAQADEXQNRRVFLLLAQ
20	orf9-1	EMIYQKWRQIEPIPGKAQKRAGWLRNVLRERGNQHLDGLEEVLAQADEGQNRRVFLLLAQ 130 140 150 160 170 180
	orf9a.pep	180 190 200 210 220 230  AAVQQDGLAQKASKAVRRAALRYEHLPEAAVADVVFSVQXREKEKAIGALQRLAKLDTEI
25	orf9-1	AAVQQDGLAQKASKAVRRAALKYEHLPEAAVADVVFSVQGREKEKAIGALQRLAKLDTEI 190 200 210 220 230 240
30	orf9a.pep	240 250 260 270 280 290  LPPTLMTLRLTARKYPEILDGFFEQTDTQNLSAVWQEMEIMNLVSLHRLDDAYARLNVLL
	orf9-1	LPPTLMTLRLTARKYPEILDGFFEQTDTQNLSAVWQEMEIMNLVSLHRLDDAYARLNVLL 250 260 270 280 290 300
35	orf9a.pep	300 310 320 330 340 350  ERNPNADLYIQAAILAANRKEXASVIDGYAEKAYGRGTGEQRGRAAMTAAMIYADRRDYT
	orf9-1	ERNPNADLYIQAAILAANRKEGASVIDGYAEKAYGRGTEEQRSRAALTAAMMYADRRDYA 310 320 330 340 350 360
40	orf9a.pep	360 370 380 390 400 410  KVRQWLKKVSAPEYLFDKGVLAAAAAVELDXGRAALRQIGRVRKLPEQQGRYFTADNLSK  !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
45	orf9-1	KVRQWLKKVSAPEYLFDKGVLAAAAAVELDGGRAALRQIGRVRKLPEQQGRYFTADNLSK 370 380 390 400 410 420
	orf9a.pep	420 430 440 450 460 470 IQMFALSKLPDKREALRGLDKIIEKPPAGSNTELQAEALVQRSVVYDRLGKRKKMISDLE
50	orf9-1	IQMLALSKLPDKREALRGLDKIIEKPPAGSNTELQAEALVQRSVVYDRLGKRKKMISDLE 430 440 450 460 470 480
•	orf9a.pep	480 490 500 510 520 530 RAFRLAPDNAQIMNNLGYSLLSDSKRLDEGFALLQTAYQINPDDTAVNDSIGWAYYLKXD
55	orf9-1	RAFRLAPDNAQIMNNLGYSLLTDSKRLDEGFALLQTAYQINPDDTAVNDSIGWAYYLKGD 490 500 510 520 530 540
60	orf9a.pep	540 550 560 570 580 590 AESALPYLRYSFENDPEPEVAAHLGEVLWALGERDQAVDVWTQAAHLTGDKKIWRETLKR
	orf9-1	AESALPYLRYSFENDPEPEVAAHLGEVLWALGERDQAVDVWTQAAHLTGDKKIWRETLKR 550 560 570 580 590 600
65	orf9a.pep	600 610 HGIALPQPSRKPRKX 
	orf9-1	HGIALPQPSRKPRKX 610

# Homology with a predicted ORF from N.gonorrhoeae

ORF9 shows 82.8% identity over a 163aa overlap with a predicted ORF (ORF9.ng) from N. gonorrhoeae:

Orf9	RFKMLTVLTATLIAGQVSAAGGGAGDMKQPKEVGKVFRKQQRYSEEEIKNERAR	54
orf9ng	MIMLPARFTILSVLAAALLAGQAYAAGAADVELPKEVGKVLRKHRRYSEEEIKNERAR	58
orf9	LAAVGERVNQIFTLLGGETALQKGQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFE	114
orf9ng		118
orf9	OAEMIYQKWRQIEPIPGKAQKRAGWLRNVLRERGNQHLDGREEVLAQADEGQ	166
orf9ng	QAEMIYQKWRQIEPIPGEAQKPAGWLRNVLKEGGNPHLDRLEEVPAQSDYVHQPMIFLLL	178
	orf9ng orf9 orf9ng orf9	orf9ng MIMLPARFTILSVLAAALLAGQAYAAGAADVELPKEVGKVLRKHRRYSEEEIKNERAR  orf9 LAAVGERVNQIFTLLGGETALQKGQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFE

The ORF9ng nucleotide sequence <SEQ ID 45> was predicted to encode a protein having including acid sequence <SEQ ID 46>:

```
20 MIMLPARFTI LSVLAAALLA GQAYAAGAAD VELPKEVGKV LRKHRRYSEE

FIKNERARLA AVGERVNRVF TLLGGETALQ KGQAGTALAT YMLMLERTKS

PEVAERALEM AVSLNAFEQA EMIYCKWRQI EPIPGEAQKP AGWLRNVLKE

EVPAQSDYVH QPMIFLLLVQ AAVQHGGVAQ KPSKAVRPAA

201 YNYEVLPETA GADAVFCVQG PQYEKAIQSF PPCGRNPQTE NIAPPFNELF

251 RPTARPISPK LLQRFFRTEP NLAKPFRPPG PEMETYQTGF PRPLTRNNPT
```

Amino acids 1-28 are a putative leader sequence, and 173-189 are predicted to be a transmembrane domain.

25 Further sequence analysis revealed the complete length ORF9ng DNA sequence <SEQ ID 47>:

	7 000 000 000 1	•				
	1	ATGTTACCCG	CCCGTTTCAC	TATTTTATCT	GTCCTCGCAG	CAGCCCTGCT
	51	TO COCCO A CAG	CCCTATCCTG	CCGGCGCGGC	GGATGTGGAG	CTGCCGAAGG
	101	<u>አአርሞሮሮርአአ</u> አ	CCTTTTAAGG	AAACATCGGC	GTTACAGCGA	GGAAGAAATC
	151	AAAAACCAAC	GCGCACGGCT	TGCGGCAGTG	GGCGAACGGG	TCAACAGGGT
30	201	CTTTTACCCTC	TTCCCCCCCTC	AAACGGCTTT	GCAGAAAGGG	CAGGCGGGAA
30	251	CCCCTCTCCC	እካርርጥልጥልጥG	CTGATGTTGG	AACGCACAAA	ATCCCCCGAA
	301	CTCCCCCAAC	CCCCCTTGGA	AATGGCCGTG	TCGCTGAACG	CGTTTGAACA
	351	CCCCCAAATC	ATTTATCAGA	AATGgcggca	gatcgagcct	ataccgggtg
	401	244444	accaGcaaaG	taactacaaa	acqtattgaa	ggaaggguua
35	451	MCACCAMC	TCCAcaaatt	granagaggTG	CtggcgcaAT	cggacgatui
22	501	CCAAAAAAcac	TTTTATEDDE	TGCTGCTGGT	GCAAGCCGCC	Gracageagg
	551	~mccccmccc	ጥሮአአአአአርሮአ	TCGAAAGCGG	TTCGCCatac	GGCGLLGanG
	601	MANGA A CAMC	TCCCCCCAAACC	aacaaTTGCC	GATGCGGTGT	TUGGUGTACA
	651	CCCACCCCAA	AACCAAAaaa	caaTCGAAGC	TTTGCAGCGT	TIGGCGMAGC
40	701	<b>ポークス サング こうごみ</b>	AATATTGCCC	CCCACTTTAA	TGACGTTGCG	TCTGACTGCA
<del>-1</del> 0	751	<b>ርርርን አ</b> ንጥአጥር	CCCAAATACT	CGACGGCTTT	TTCGAGCAGA	CAGACACCCCA
	801	አአአሮሮመመምሮር	GCCGTCTGGC	AGGAAATGGA	AATTATGAAT	CTGGTTTCCC
	851	<b>からくとのですることとこ</b>	CCATCATCCC	TATGCGCGTT	TGAACGTGCT	GTTGGAACAC
	901	እ አ <i>ርርርር</i> እስጥር	CAAACCTGTA	TATTCAGGCG	GCGATATTGG	CGGCAAACCG
45	951	AAAAGAAGGT	GCGTCCGTTA	TCGACGGCTA	CGCCGAAAAG	GCATACGGCA
73	1001	GGGGGACGGG	GGAACAGCGG	GGCagggcgg	cAATgacggc	GGCGATGATA
	1051	mamccccacc	. CCDCCCDTTD	CGCCAAAGTC	AGGCAGTGGT	IGHMAMAGI
	1101	GTCCGCGCCG	GAATACCTGT	TCGACAAAGG	CGTGCTGGCG	GCTGCGCGG
	1151	ርጥርርርር እስጥፕ	GGACGGAGGC	CGGGCGGCTT	TGCGGCAGAT	CGGCAGGGTG
50	1201	CGGAAACTTC	CCGAACAGCA	GGGGCGGTAT	TTTACGGCAG	ACAATTIGIC
50	1251	CAAAATACAG	ATGCTCGCCC	TGTCGAAGCT	GCCCGACAAA	CGGGAAGCCC
	1301	MCAMCCCCCT	GAACAACATC	: ATCGCCAAAC	TTTCGGCGGC	GGGAAGCACG
	135	GAACCTTTGG	G CGGAAGCATT	GGCACAGCGT	TCCATTATTT	ACGaacAGTT
	140:		GGAAAAATGA	A TTGCCGACCT	tgaAACcgcg	CTCAAACTTA
55	145	CCCCCCATA	<b>ላ ጥርሮልሮልልልጥ</b> ባ	' ATGAATAATC	TGGGCTACAG	CCTGCTTTCC
33	150	- CAMMCCAAAA	C GTTTGGACGA	A GGGTTTCGCC	CTGCTTCAGA	CGGCATACCA
	155	1 XXTCXXCCC	CACGATACCO	: CCGTTAACGA	CAGCATAGGC	TGGGGGIAII
	160	***********	- CCDCacaaal	A AGCGCGCTGC	CGTATCTGCG	gtattegttt
	165	1 gAAAACGAC	C CCGAGCCCG	A AGTTGCCGCC	CATTTGGGCG	AAGTGTTGTG
		-				

	•	•						
	1751 AC	CTTAGGGG	AGACAAGAAA	ATATGGCG	GG AGACGCT	SACG CAGGCG		
			CCGAGCCTTC					
	This encodes a pro-	tein navin	g amino aci	a sequence	e <2FO m	48>:		:
5	51 KN 101 VA	ERARLAAV ERALEMAV	GERVNRVFTI SLNAFEQAEN	LGGETALO	KG QAGTALA EP IPGEAQE	CVLR KHRRYS ATYM LMLERT CPAG WLRNVI	KSPE KEGG	
						AQKA SKAVRR EILP PTLMTI		
10	251 RF 301 NF 351 YF	CYPEILDGF PNANLYIQA ADRRDYAKV	FEOTDTONLS AILAANRKEO ROWLKKVSAE	S AVWQEMEI S ASVIDGYA P EYLFDKGV	MN LVSLRKI AEK AYGRGTO YLA AAAAAEI	PDDA YARLNY SEQR GRAAMT LDGG RAALRO	'LLEH 'AAMI )IGRV	
			-			LNNI IAKLSA NAQI MNNLGY		
15	551 EN		HLGEVLWALO			DAE SALPYI DKK IWRETI		
	ORF9ng and ORF9			ity in 614	aa overlap:			
			10	20	30	40	50	60
20	orf9-1.pep		MLTVLTATLIA	AGQVSAAGGG	SAGDMKQPKEV	/GKVFRKQQRY	SEEEIKNER	ARLA
	orf9ng-1					:  ::   /GKVLRKHRRY		
	•		10	20	30	40	50	
25	orf0-1 non	AUCEDUM	70	80 NT OVCONCE?	90	100 RTKSPEVAERA	110	120
	orf9-1.pep	1111111	::!!!!!!!!	1111111111			1111111111	HĪL
-	orf9ng-1	AVGERVNI 60	RVFTLLGGET <i>i</i> 70	ALQKGQAGT <i>I</i> 80	LATYMLMLEI 90	RTKSPEVAERA 100	ALEMAVSLNA 110	FEQA
30	•		130 :	L40	150	160	170	180
	orf9-1.pep	EMIYQKW	RQIEPIPGKA	QKRAGWLRNV	/LRERGNQHLI	OGLEEVLAQAI	DEGQNRRVFL	LLAQ
	orf9ng-1	EMIYQKW	RQIEPIPGEA(	OKPAGWLRN'	/LKEGGNQHLI	:     :   GLKEVLAQSI		
35	· .	120	130	140	150	. 160	170	
	orf9-1.pep			200 RAAT.KYEHT.	210 PEAAVADVVE	220 SVQGREKEKAI	230	240 DTET
40			:1111111111	[11][1][1][1]		:111111111	шінн	1111
40	orf9ng-1	180	VAQKASKAVRI 190	200	210	GVQGREKEKA] 220	230	DIEI
			250 2	260	270	280	290	300
45	orf9-1.pep	LPPTLMT	LRLTARKYPE:	ILDGFFEQTI	TQNLSAVWQ	EMEIMNLVSLF	IRLDDAYARI	NVLL
73	orf9ng-1	LPPTLMT	LRLTARKYPE:	ILDGFFEQTI	DTQNLSAVWQI	EMEIMNLVSLE	RKPDDAYARI	NVLL
		240	250	260	270	280	290	
50	orf9-1.pep			320 NRKEGASVII	330 GYAEKAYGRO	340 Steeorsraai	350 LTAAMMYADR	360 RDYA
	orf9ng-1	1:1111:	ПППП	шшш			:1111:1111	HH
	OII 3Hg~I	300	310	320	330	340	350	RDIA
55	•		370 :	380	390	400	410	420
	orf9-1.pep					RQIGRVRKLPH		
	orf9ng-1	KVRQWLK	KVSAPEYLFDI	KGVLAAAAA	AELDGGRAAL	RQIGRVRKLP	EQQGRYFTAL	NLSK
60		360	370	380	390	400	410	
	orf9-1.pep			440 GLDKIIEKPI	450 PAGSNTELOA	460 Ealvorsvvyi	470 DRLGKRKKMI	480 ISDLE
	• •	1111111	1111111	11::11	1:::11	:   :: : EALAQRSIIY	:::111 111	:111
65	orf9ng-l	1QMLALS 420	430	440	450	£ALAQRS11YI 460	EQFGKRGKMI 470	.AULE
			490	500	510	520	530	540

```
RAFRLAPDNAQIMNNLGYSLLTDSKRLDEGFALLQTAYQINPDDTAVNDSIGWAYYLKGD
                     orf9-1.pep
                     TALKLTPDNAQIMNNLGYSLLSDSKRLDEGFALLQTAYQINPDDTAVNDSIGWAYYLKGD
         orf9ng-1
                                                                530
                                                       520
                                     500
                    480
                            490
5
                                            570
                                                     580
                     AESALPYLRYSFENDPEPEVAAHLGEVLWALGERDQAVDVWTQAAHLTGDKKIWRETLKR
                                    560
                     пининининининининий прининий прининий
         orf9-1.pep
                     AESALPYLRYSFENDPEPEVAAHLGEVLWALGERDQAVDVWTQAAHLRGDKKIWRETLKR
         orf9ng-1
                                              570
                                                      580
                                      560
                             550
                    540
10
                           610
                     HGIALPQPSRKPRKX
         orf9-1.pep
                     YGIALPEPSRKPRKX
15
         orf9ng-1
                             610
                    600
     In addition, ORF9ng shows significant homology with a hypothetical protein from P.aeruginosa:
          sp|P42810|YHE3_PSEAE HYPOTHETICAL 64.8 KD PROTEIN IN HEMM-HEMA INTERGENIC REGION
          >gi|1072999|pir||S49376 hypothetical protein 3 - Pseudomonas aeruginosa >gi|557259
          (X82071) orf3 [Pseudomonas aeruginosa] Length = 576
20
           Score = 128 bits (318), Expect = 1e-28
           Identities = 138/587 (23%), Positives = 228/587 (38%), Gaps = 125/587 (21%)
          Query: 67 VFTLLGGETALQKGQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFEQAEMIYQKWR 126
          +++LL E A Q+ + AL+ Y++ ++T+ P V+ERA +A L A ++A W
Sbjct: 53 LYSLLVAELAGQRNRFDIALSNYVVQAQKTRDPGVSERAFRIAEYLGADQEALDTSLLWA 112
25
          Query: 127 QIEPIPGEAQKPAG------WLRNVLKEGGNQHLDGLKEVLAQSDDVQKRRI 172
                                  ++ VL G+ H D L
                    + P +AQ+ A
          Sbjct: 113 RSAPDNLDAQRAAAIQLARAGRYEESMVYMEKVLNGQGDTHFDFLALSAAETDPDTRAGL 172
 30
          KY + + A+ Q ++A+ L+ +
                                     ++
          Sbjct: 173 L-----QSFDHLLKKYPNNGQLLFGKALLLQQDGRPDEALTLLEDNS 214
 35
           Query: 233 KLDTEILPPTLMTLRLTARK-----YPEILDGFFEQTDTQNLSAVWQEMEIMNLVSLRKP 287
                                          P + G E D + + + + LV
                        E+ P L + L + K
           Sbjct: 215 ASRHEVAPLLLRSRLLQSMKRSDEALPLLKAGIKEHPDDKRVRLAYARL---LVEQNRL 270
           Query: 288 DDAYARLNVLLEHNPN----- 312
 40
                                                    A +Y++ +
                            L++ P+
           Sbjct: 271 DDAKAEFAGLVQQFPDDDDDLRFSLALVCLEAQAWDEARIYLEELVERDSHVDAAHFNLG 330
           Query: 313 -LAANRKEGASVIDGYAEKAYGRGTGEQRGRAAMTAAMIYADRRDYAKVRQWLKKVSAPE 371
  45
                                              + T ++ A R D A R + P+
                      LA +K+ A +D YA+ G G
           Sbjct: 331 RLAEEQKDTARALDEYAQ--VGPGNDFLPAQLRQTDVLLKAGRVDEAAQRLDKARSEQPD 388
           A L I+ ALS
           Y A L I+ ALS +
Sbjct: 389 Y-----AIQLYLIEAEALSNNDQQE 408
  50
           Query: 432 EALIGLNNIIAKLSAAGSTEPLAEALAQRSIIYEQFGKRGKMIADLETALKLTPDNAQIM 491
                     +A + + + E L L RS++ E+ +M DL + PDNA +
            Sbjct: 409 KAWQAIQEGLKQYP----EDL-NLLYTRSMLAEKRNDLAQMEKDLRFVIAREPDNAMAL 462
  55
            Query: 492 NNLGYSLLSDSKRLDEGFALLQTAYQINPDDTAVNDSIGWAYYLKGDAESALPYLRYSFE 551
                     N LGY+L + R E L+ A+++NPDD A+ DS+GW Y +G A YLR + +
            Sbjct: 463 NALGYTLADRTTRYGEARELILKAHKLNPDDPAILDSMGWINYRQGKLADAERYLRQALQ 522
            Query: 552 NDPEPEVAAHLGEVLWALGERDQAVDVWTQAAHLRGDKKIWRETLKR 598
  60
                       P+ EVAAHLGEVLWA G + A +W + + D + R T+KR
            Sbjct: 523 RYPDHEVAAHLGEVLWAOGROGDARAIWREYLDKOPDSDVLRRTIKR 569
             gi|2983399 (AE000710) hypothetical protein [Aquifex aeolicus] Length = 545
   65
             Score = 81.5 bits (198), Expect = 1e-14
             Identities = 61/198 (30%), Positives = 98/198 (48%), Gaps = 19/198 (9%)
            Query: 408 GRYFTADNL-SKIQMLALSKLPDKREALIGLNNIIAKLSAAGSTEPLAEALAQ----- 459
                                        PDK+E L + +K
                      GYALK++LA
   70
```

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 7

The following partial DNA sequence was identified in N. meningitidis <SEQ ID 49>:

```
AACCTCTACG CCGGCCCGCA GACCACATCC GTCATCGCAA ACATCGCCGA
20
                51
                    CAACCTGCAA CTGGCCAAAG ACTACGGCAA AGTACACTGG TTCGCCTCCC
               101
                    CGCTCTTCTG GCTCCTGAAC CAACTGCACA ACATCATCGG CAACTGGGGC
                    TGGGCGATTA TCGTTTTAAC CATCATCGTC AAAGCCGTAC TGTATCCATT
               151
                    GACCAACGCC TCTTACCGCT CTATGGCGAA AATGCGTGCC GCCGCACCCA
               201
                    AACTGCAAGC CATCAAAGAG AAATACGGCG ACGACCGTAT GGCGCAACAA
               251
25
               301
                    CAGGCGATGA TGCAGCTTTA CACAGACGAG AAAATCAACC CGACTGGGCG
                    GCTGCCTGCC TATGCTGTTG CAAATCCCCG TCTTCATCGG ATTGTATTGG
               351
                    GCATTGTTCG CCTCCGTAGA ATTGCGCCAG GCACCTTGGC TGGGTTGGAT
                401
                    TACCGACCTC AGCCGCGCG ACCCCTACTA CATCCTGCCC ATCATTATGG
                451
                501
                    CGGCAACGAT GTTCGCCCAA ACTTATCTGA ACCCGCCGCC GACCGACCCG
30
                551
                    ATGCagGCGA AAATGATGAA AATCATGCCG TTGGTTTTCT CsGwCrTGTT
                    CTTCTTCTTC CCTGCCGGks TGGTATTGTA CTGGGTAGTC AACAACCTCC
                601
                651 TGACCATCGC CCAGCAATGG CACATCAACC GCAGCATCGA AAAACAACGC
                701 GCCCAAGGCG AAGTCGTTTC CTAA
```

This corresponds to the amino acid sequence <SEQ ID 50; ORF11>:

35	1	NLYAGPQTTS	VIANIADNLQ	LAKDYGKVHW	FASPLFWLLN	QLHNIIGNWG
	51	WAIIVLTIIV	KAVLYPLTNA	SYRSMAKMRA	AAPKLQAIKE	KYGDDRMAQQ
	101	QAMMQLYTDE	KINPLGGCLP	MLLQIPVFIG	LYWALFASVE	LRQAPWLGWI
	151	TDLSRADPYY	ILPIIMAATM	FAQTYLNPPP	TDPMQAKMMK	IMPLVFSXXF
	201	FFFPAGXVLY	WVVNNLLTIA	QQWHINRSIE	KQRAQGEVVS	*

40 Further sequence analysis revealed the complete DNA sequence <SEQ ID 51>:

	1	ATGGATTTTA	AAAGACTCAC	GGCGTTTTTC	GCCATCGCGC	TGGTGATTAT
	51	GATCGGCTGG	GAAAAGATGT	TCCCCACTCC	GAAGCCAGTC	CCCGCGCCCC
	101	AACAGGCAGC	ACAACAACAG	GCCGTAACCG	CTTCCGCCGA	AGCCGCGCTC
	151	GCGCCCGCAA	CGCCGATTAC	CGTAACGACC	GACACGGTTC	AAGCCGTCAT
45	201	TGATGAAAAA	AGCGGCGACC	TGCGCCGGCT	GACCCTGCTC	AAATACAAAG
	251	CAACCGGCGA	CGAAAATAAA	CCGTTCATCC	TGTTTGGCGA	CGGCAAAGAA
•	301	TACACCTACG	TCGCCCAATC	CGAACTTTTG	GACGCGCAGG	GCAACAACAT
	351	TCTAAAAGGC	ATCGGCTTTA	GCGCACCGAA	AAAACAGTAC	AGCTTGGAAG
	401	GCGACAAAGT	TGAAGTCCGC	CTGAGCGCGC	CTGAAACACG	CGGTCTGAAA
50	451	ATCGACAAAG	TTTATACTTT	CACCAAAGGC	<b>AGCTATCTGG</b>	TCAACGTCCG
	501	CTTCGACATC	GCCAACGGCA	GCGGTCAAAC	CGCCAACCTG	AGCGCGGACT
	551	ACCGCATCGT	CCGCGACCAC	AGCGAACCCG	AGGGTCAAGG	TTACTTTACC
	601	CACTCTTACG	TCGGCCCTGT	TGTTTATACC	CCTGAAGGCA	ACTTCCAAAA
	651	AGTCAGCTTT	TCCGACTTGG	ACGACGATGC	CAAATCCGGC	AAATCCGAGG
55	701	CCGAATACAT	CCGCAAAACC	CCGACCGGCT	GGCTCGGCAT	GATTGAACAC
	751	CACTTCATGT	CCACCTGGAT	TCTCCAACCT	AAAGGCAGAC	AAAGCGTTTG
	801	CGCCGCAGGC	GAGTGCAACA	TCGACATCAA	ACGCCGCAAC	GACAAGCTGT
	851	ACAGCACCAG	CGTCAGCGTG	CCTTTAGCCG	CCATCCAAAA	CGGCGCGAAA
	901	GCCGAAGCCT	CCATCAACCT	CTACGCCGGC	CCGCAGACCA	CATCCGTCAT
60	951	CGCAAACATC	GCCGACAACC	TGCAACTGGC	CAAAGACTAC	GGCAAAGTAC

	1001 1051 1101 1151	ATCGGCAACT CGTACTGTAT GTGCCGCCGC	CTCCCCGCTC GGGGCTGGGC CCATTGACCA ACCCAAACTG	GATTATCGTT ACGCCTCTTA CAAGCCATCA	TTAACCATCA CCGCTCTATG AAGAGAAATA	TCGTCAAAGC GCGAAAATGC CGGCGACGAC
5	1201 1251 1301 1351 1401	CAACCCGCTG TCGGATTGTA TGGCTGGGTT GCCCATCATT	AACAACAGGC GGCGGCTGCC TTGGGCATTG GGATTACCGA ATGGCGGCAA	TGCCTATGCT TTCGCCTCCG CCTCAGCCGC CGATGTTCGC	GTTGCAAATC TAGAATTGCG GCCGACCCCT CCAAACTTAT	CCCGTCTTCA CCAGGCACCT ACTACATCCT CTGAACCCGC
10	1451 1501 1551 1601	TTCTCCGTCA AGTCAACAAC	CCCGATGCAG TGTTCTTCTT CTCCTGACCA ACGCGCCCAA	CTTCCCTGCC TCGCCCAGCA	GGTCTGGTAT ATGGCACATC	TGTACTGGGT

This corresponds to the amino acid sequence <SEQ ID 52; ORF11-1>:

```
MDFKRLTAFF AIALVIMIGW EKMFPTPKPV PAPQQAAQQQ AVTASAEAAL
15
                     APATPITVTT DTVQAVIDEK SGDLRRLTLL KYKATGDENK PFILFGDGKE
                 51
                     YTYVAQSELL DAQGNNILKG IGFSAPKKQY SLEGDKVEVR LSAPETRGLK
                101
                     IDKVYTFTKG SYLVNVRFDI ANGSGQTANL SADYRIVRDH SEPEGQGYFT
                151
                     HSYVGPVVYT PEGNFQKVSF SDLDDDAKSG KSEAEYIRKT PTGWLGMIEH
                201
                     HFMSTWILQP KGRQSVCAAG ECNIDIKRRN DKLYSTSVSV PLAAIQNGAK
20
                251
                     AEASINLYAG PQTTSVIANI ADNLQLAKDY GKVHWFASPL FWLLNQLHNI
                301
                     IGNWGWAIIV LTIIVKAVLY PLTNASYRSM AKMRAAAPKL QAIKEKYGDD
                351
                     RMAQQQAMMQ LYTDEKINPL GGCLPMLLQI PVFIGLYWAL FASVELRQAP
                401
                     WLGWITDLSR ADPYYILPII MAATMFAQTY LNPPPTDPMQ AKMMKIMPLV
                451
                     FSVMFFFFPA GLVLYWVVNN LLTIAQQWHI NRSIEKQRAQ GEVVS*
25
                501
```

Computer analysis of this amino acid sequence gave the following results:

Homology with a 60kDa inner-membrane protein (accession P25754) of *Pseudomonas putida* ORF11 and the 60kDa protein show 58% aa identity in 229 aa overlap (BLASTp).

```
LYAGPQTTSVIANIADNLQLAKDYGKVHWFASPLFWLLNQLHNIIGNWGWAIIVLTIIVK 61
          ORF11
                     LYAGP+ S + ++ L+L DYG + + A P+FWLL +H+++GNWGW+IIVLT+++K
30
                 324 LYAGPKIQSKLKELSPGLELTVDYGFLWFIAQPIFWLLQHIHSLLGNWGWSIIVLTMLIK 383
          60K
                     AVLYPLTNASYRSMAKMRAAAPKLQAIKEKYGDDRXXXXXXXXXLYTDEKINPLGGCLPM 121
          ORF11
                                                                 LY EKINPLGGCLP+
                       + +PL+ ASYRSMA+MRA APKL A+KE++GDDR
                 384 GLFFPLSAASYRSMARMRAVAPKLAALKERFGDDRQKMSQAMMELYKKEKINPLGGCLPI 443
35
           60K
                 122 LLQIPVFIGLYWALFASVELRQAPWLGWITDLSRADPYYILPIIMAATMFAQTYLNPPPT 181
           ORF11
                      L+Q+PVF+ LYW L SVE+RQAPW+ WITDLS DP++ILPIIM ATMF Q LNP P
                  444 LVQMPVFLALYWVLLESVEMRQAPWILWITDLSIKDPFFILPIIMGATMFIQQRLNPTPP 503
           60K
40
                 182 DPMQAKMMKIMPLVXXXXXXXPAGXVLYWVVNNLLTIAQQWHINRSIE 230
           ORF11
                                            PAG VLYWVVNN L+I+QQW+I R IE
                      DPMQAK+MK+MP++
                  504 DPMQAKVMKMMPIIFTFFFLWFPAGLVLYWVVNNCLSISQQWYITRRIE 552
           60K
```

Homology with a predicted ORF from N.meningitidis (strain A)

ORF11 shows 97.9% identity over a 240aa overlap with an ORF (ORF11a) from strain A of N.

meningitidis:

						10	20	30
	orfl1.pep				NLYAGPQ'		NLQLAKDYGI	KVHW
50	• •				1111111	1111111111		1111
30	orflla	IKRRNDKI	LYSTSVSVPI	AAIQNGAKSX	<b>ASINLYAGPQ</b>	TTSVIANIAL	NLQLXKDYG	KVHW
	OLLILA	280	290	300	310	320	330	
			40	50	60	70	80	90
55	orfl1.pep	FASPLEW	LLNOLHNII	ENWGWAIIVLT	IIVKAVLYPL	TNASYRSMAH	MRAAAPKLQ:	AIKE
33	Ollil.pcp	1111111	111111111		1111111111	111111111		1111
	orf11a	FASPLFW	LLNOLHNII 350	NWGWAIIVLT 360	IIVKAVLYPL 370	TNASYRSMAI 380	MRAAAPKLQ 390	AIKE

	orfl1.pep		00 11 .QQQAMMQLYTD				150	
	offif.pep							
_	orflla	KYGDDRMA	QQQAMMQLYTD	EKINPLGGCLE	MLLQIPVFIG	LYWALFASVE	LRQAPWLGWI	
5		400	410	420	430	440	450	
		1	60 17	0 180	190	200	210	
	orfl1.pep		YYILPIIMAAT					
10	orflla		 TAAMIITIIIYY					
IU	OLLIIA	460	470	480	490	500	510	
	•							
	611		20 23			•		
15	orf11.pep		'IAQQWHINRSI 					
10	orflla ,		IAQQWHINRSI			,		
		520	530	540		-		
	The complete les	noth ORF11a	nucleotide :	sequence <s< td=""><td>EO ID 53&gt;</td><td>is:</td><td></td><td></td></s<>	EO ID 53>	is:		
	The complete ter		· naoroonao	ocquestoo 35	24 10 00	10.		
	1	ANGGATTTTA	AAAGACTCAC	NGNGTTTTTC	GCCATCGCAC	TGGTGATTA	T	
20	51	GATCGGATNG					-	
	101	AACAGACGGC						
	151 201	GCGCCCGNAN TGATGAAAAA						
	251	CAACCGGCGA						
25	301	TACACCTACN						
	351		ATCGGCTTTA					
	401		TGAAGTCCGC					
	451	ATCGACAAAG						
20	501	CTTCGACATC						
30	551 601	ACCGCATCGT CACTCTTACG						
	651		TCCGACTTGG					
	701		CCGCAAAACC					
	751		CCACCTGGAT					
35	801		GACTGCNGTA					
	851		CGTCAGCGTG					
	.901		CCATCAACCT					
	951 1001	ACTGGTTCGC	GCCGACAACC					
40	1051		GGGGCTGGGC				_	•
70	1101		CCATTGACCA					
	1151		GCCCAAACTG				-	
	1201		AGCAACAAGC				_	
4.5	1251		GGCGGCTGCC					
45	1301		TTGGGCATTG					
	1351 1401		GGATTACCGA ATGGCGGCAA					
	1451		CCCGATGCAG				-	
	1501		NGTTCTTCNN					
50	1551		CTCCTGACCA					
	1601	TCGAAAAACA	ACGCGCCCAA	GGCGAAGTCG	TTTCCTAA			
	This encodes a p	rotein havin	a amino acid	l seguence <	SEO ID 545	»·		
	Tims checodes a p	protein mavin	g animo acio	soquence \	ord in 24%	•		
	1	XDFKRLTXFF	AIALVIMIGX	XXMFPTPKPV	PAPOOTAOOC	AVXASAEAA	L	
	51		DTVQAVIDEK					
55	101	YTYXAXSELL	DAQGNNILKG	IGFSAPKKQY	SLEGDKVEVE	LSAPETRGL	K	
	151		SYLVNVRFDI			_		
	201		PEGNFQKVSF					
	251		KGGQSVCAAG					
60	301 351		POTTSVIANI					
UU	401		LTIIVKAVLY LYTDEKINPL					
	451		ADPYYILPII					
	501		GLVLYWVINN				_	
	ODE112 and OT							
	ORF11a and OI	TII-I SHOW	33.270 IGCN	лгу ш <i>э</i> 44 аа	overiap:			
65			10	20	30	40	50 .	60
05			10	20	50	-10		,0

	orflla.pep	XDFKRLTXFFAIALVI			: ! ! ! ! ! ! ! . !		
	orf11-1		MIGWEKMF	PTPKPVPAPQQ 30	AAQQQAVTA: 40	SAEAALAPATE 50	PITVTT 60
5		70	80	90	100	110	120
	orflla.pep	DTVQAVIDEKSGDLRR:		111 111111			, , , , , ,
10	orf11-1	DTVQAVIDEKSGDLRR 70	80	90	100	110	120
		130 IGFSAPKKQYSLEGDK	140 VEVRLSAE	150 ETRGLKIDKV	160 TFTKGSYLV	170 NVRFDIANGS	180 GOTANL
15	orflla.pep			PETRGLKIDKV	(TFTKGSYLV	NVRFDIANGS	11111
15	OIIII I	130	140	150 210	160 220	170 230	240
	orflla.pep	190 SADYRIVRDHSEPEGÇ	111111	VGPVVYTPEGN	FOKVSFSDLD	DDAXSGKSEA	EYIRKT
20	orf11-1	SADYRIVRDHSEPEGQ	GYFTHSY 200	VGPVVYTPEGN 210	FOKVSFSDLI 220	DDAKSGKSEA 230	EYIRKT 240
		050	2,60	270	280	290	300
25	orflla.pep	XTGWLGMIEHHFMSTV             PTGWLGMIEHHFMSTV		11111111	FILL 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1   1   1   1   1   1   1	
	orf11-1	PTGWLGMIEHHFMSTV 250	260 260	270	280	290	,
30		310 SXASINLYAGPQTTS	320 VIANIADN	330 ILQLXKDYGKVH	340 WFASPLFWL	350 LNQLHNIIGN	360 WGWAIIV
	orflla.pep	SXASINLYAGPQITS :              AEASINLYAGPQTTS	 VIANIADN	ILQLAKDYGKVH			,,,,,,,
35		310	320 380	330 390	400	410	420
	orflla.pep	370 LTIIVKAVLYPLTNA 	SYRSMAK	MRAAAPKLQAII	KEKYGDDRMA	1111111111	111111
40	orf11-1	LTIIVKAVLYPLTNA	SYRSMAKN 380	MRAAAPKLQAII 390	KEKYGDDRMA 400	QQQAMMQLYT 410	DEKINPL 420
40		430 GGCLPMLLQIPVFIG	440	450	460	470	480 TMFAOTY
	orf11a.pep	GGCLPMLLQIPVFIG 			1 1 1 1 1 1 1 1 1 1 1		TMFAQTY
45	orf11-1	430	440	450	460	470	100
	orflla.pep	490 LNPPPTDPMQAKMMI	500 KIMPLVXS	510 XXFFXFPAGLV	520 LYWVINNLL	530 PIAQQWHINRS	540 SIEKQRAQ
50	orf11-1		 KIMPLVFS				
		490	500	310	320		
55	orflla.pep	GEVVSX					
	orfll-1	GEVVSX					

# 60 Homology with a predicted ORF from N.gonorrhoeae

ORF11 shows 96.3% identity over a 240aa overlap with a predicted ORF (ORF11.ng) from N. gonorrhoeae:

	a . 511	NLYAGPOTTSVIANIADNLOLAKDYGKVHWFASPLFWLLNOLHNIIGNWGWAIIVLT	57
	Orf11		
65	orfllng	MAVNLYAGPQTTSVIANIADNLQLAKDYGKVHWFASPLFWLLNQLHNIIGNWGWAIVVLT	00

	orf11	IIVKAVLYPLTNASYRSMAKMRAAAPKLQAIKEKYGDDRMAQQQAMMQLYTDEKINPLGG	117
	orfllng	IIVKAVLYPLTNASYRSMAKMRAAAPELQTIKEKYGDDRMAQQQAMMQLFEDEEINPLGG	120
5	orf11	CLPMLLQIPVFIGLYWALFASVELRQAPWLGWITDLSRADPYYILPIIMAATMFAQTYLN	177
	orf11ng	CLPMLLQIPVFIGLYWALFASVELRQAPWLGWITDLSRADPYYILPIIMAATMFAQTYLN	180
10	orf11	PPPTDPMQAKMMKIMPLVFSXXFFFFPAGXVLYWVVNNLLTIAQQWHINRSIEKQRAQGE	237
10	orfllng	PPPTDPMQAKMMKIMPLVFSVMFFFFPAGLVLYWVVNNLLTIAQQWHINRSIEKQRAQGE	240
	orf11	VVS 240	
15	orfllng	VVS 243	

An ORF11ng nucleotide sequence <SEQ ID 55> was predicted to encode a protein having amino acid sequence <SEQ ID 56>:

```
1 MAVNLYAGPQ TTSVIANIAD NLQLAKDYGK VHWFASPLFW LLNQLHNIIG
51 NWGWAIVVLT IIVKAVLYPL TNASYRSMAK MRAAAPELQT IKEKYGDDRM
20 101 AQQQAMMQLF EDEEINPLGG CLPMLLQIPV FIGLYWALFA SVELRQAPWL
151 GWITDLSRAD PYYILPIIMA ATMFAQTYLN PPPTDPMQAK MMKIMPLVFS
201 VMFFFFPAGL VLYWVVNNLL TIAQQWHINR SIEKQRAQGE VVS*
```

Further sequence analysis revealed the complete gonococcal DNA sequence <SEQ ID 57> to be:

```
ATGGATTTTA AAAGACTCAC GGCGTTTTTC GCCATCGCGC TGGTGATTAT
25
                     GATCGGCTGG GAAAAAATGT TCCCCACCCC GAAACCCGTC CCCGCGCCCC
               101
                    AACAGGCGGC ACAAAAACAG GCAGCAACCG CTTCCGCCGA AGCCGCGCTC
                     GCGCCCGCAA CGCCGATTAC CGTAACGACC GACACGGTTC AAGCCGTTAT
                     TGATGAAAAA AGTGGCGACC TGCGCCGGCT GACCCTGCTC AAATACAAAG
               201
                     CAACCGGCGA CGAAAACAAA CCGTTCGTCC TGTTTGGCGA CGGCAAAGAA
               251
30
                301
                     TACACCTACG TCGCCCAATC CGAACTTTTG GACGCGCAGG GCAACAACAT
                     TCTGAAAGGC ATCGGCTTTA GCGCACCGAA AAAACAGTAC ACCCTCAACG
                351
                     GCGACACAGT CGAAGTCCGC CTGAGCGCGC CCGAAACCAA CGGACTGAAA
                401
                     ATCGACAAAG TCTATACCTT TACCAAAGAC AGCTATCTGG TCAACGTCCG
                451
                501
                    CTTCGACATC GCCAACGGCA GCGGTCAAAC CGCCAACCTG AGCGCGGACT
35
                551
                     ACCGCATCGT CCGCGACCAC AGCGAACCCG AGGGTCAAGG CTACTTTACC
                601
                     CACTCTTACG TCGGCCCTGT TGTTTATACC CCTGAAGGCA ACTTCCAAAA
                     AGTCAGCTTC TCCgacTTgg acgACGATGC gaaaTccggc aaATccgagg
                     ccgaatacaT CCGCAAAACC ccgaccggtt ggctcggcat gattgaacac
                701
                751
                     cacttcatgt ccacctggat cctccAAcct aaaggcggcc aaaacgtttg
40
                801
                    cgcccaggga gactgccgta tcgacattaa aCgccgcaac gacaagctgt
                851
                     acagcgcaag cgtcagcgtg cctttaaccg ctatcccaac ccgggggcca
                901
                     aaaccgaaaa tggcggTCAA CCTGTATGCC GGTCCGCAAA CCACATCCGT
                     TATCGCAAAC ATCGCcgacA ACCTGCAACT GGCAAAAGAC TACGGTAAAG
                951
               1001
                     TACACTGGTT CGCATCGCCG CTCTTCTGGC TCCTGAACCA ACTGCACAAC
45
                     ATTATCGGCA ACTGGGGCTG GGCAATCGTC GTTTTGACCA TCATCGTCAA
               1051
               1101
                     AGCCGTACTG TATCCATTGA CCAACGCCtc ctACCGTTCG ATGGCGAAAA
               1151
                     TGCGTGccgc cgcacCcaaA CTGCAGACCA TCAAAGAAAA ATAcgGCGAC
                     GACCGTATGG CGCAACAGCA AGCGATGATG CAGCTTTACA AAGACGAGAA
               1201
               1251
                     AATCAACCCG CTGGGCGGCT GTctgcctat gctgttgCAA ATCCCCGTCT
50
                     TCATCGGCTT GTACTGGGCA TTGTTCGCCT CCGTAGAATT GCGCCAGGCA
               1301
                     CCTTGGCTGG GCTGGATTAC CGACCTCAGC CGCGCCGACC CCTACTACAT
               1351
                     CCTGCCCATC ATTATGGCGG CAACGATGTT CGCCCAAACC TATCTGAACC
               1401
               1451
                     CGCCGCCGAC CGACCCGATG CAGGCGAAAA TGATGAAAAT CATGCCGTTG
               1501
                     GTTTTCTCCG TCATGTTCTT CTTCTTCCCT GCCGGTTTGG TTCTCTACTG
55
               1551
                     GGTGGTCAAC AACCTCCTGA CCATCGCCCA GCAGTGGCAC ATCAACCGCA
               1601
                     GCATCGAAAA ACAACGCGCC CAAGGCGAAG TCGTTTCCTA A
```

This encodes a protein having amino acid sequence <SEQ ID 58; ORF11ng-1>:

,	1	MDFKRLTAFF	AIALVIMIGW	EKMFPTPKPV	PAPQQAAQKQ	AATASAEAAL
	51	APATPITVTT	DTVQAVIDEK	SGDLRRLTLL	KYKATGDENK	PFVLFGDGKE
60	. 101	YTYVAQSELL	DAQGNNILKG	IGFSAPKKQY	TLNGDTVEVR	LSAPETNGLK
	151	IDKVYTFTKD	SYLVNVRFDI	ANGSGQTANL	SADYRIVRDH	SEPEGOGYFT
	201	HSYVGPVVYT	PEGNFQKVSF	SDLDDDAKSG	KSEAEYIRKT	PTGWLGMIEH
	251	HFMSTWILQP	KGGQNVCAQG	DCRIDIKRRN	DKLYSASVSV	PLTAIPTRGP
	301	KPKMAVNLYA	GPQTTSVIAN	IADNLQLAKD	YGKVHWFASP	LFWLLNQLHN

351 IIGNWGWAIV VLTIIVKAVL YPLTNASYRS MAKMRAAAPK LQTIKEKYGD 401 DRMAQQQAMM QLYKDEKINP LGGCLPMLLQ IPVFIGLYWA LFASVELRQA 451 PWLGWITDLS RADPYYLLPI IMAATMFAQT YLNPPPTDDM QAKMKIMPL

451

VFSVMFFFFP AGLVLYWVVN NLLTIAQQWH INRSIEKQRA QGEVVS\*

ORF11ng-1 and ORF11-1 shown 95.1% identity in 546 aa overlap: 5

5	ORFIIng-1 and ORF	11-1 SHOWII 93.1	70 Identity i	11 540 44 01	ortup.		
	orfllng-l.pep	10 MDFKRLTAFFAIAL	20 VIMIGWEKMF	30 PTPKPVPAPQ	40 QAAQKQAATAS	50 BAEAALAPATI	60 TTVTT
	Office of the contract of the		1111111111	:	1111:11:11		[
10	orf11-1	MDFKRLTAFFAIAL 10	VIMIGWEKMF 20	30	40	50	60
	orfllng-1.pep	70 DTVQAVIDEKSGDL	80 RRLTLLKYKA	90 TGDENKP <b>FV</b> L	100 FGDGKEYTYV	110 AQSELLDAQGI	120 NNILKG
15	orfl1-1		 RRLTLLKYKA 80	 TGDENKPFIL 90	FGDGKEYTYV 100		NNILKG 120
		130	140	150	160	170	180
	orfllng-1.pep	TOFSAPKKOYTING	DTVEVRLSAP	ETNGLKIDKV	YTFTKDSYLV	NVRFDIANGS	GOTANL
20	orfll-1		 DKVEVRLSAP 140	ETRGLKIDKV	YTFTKGSYLV 160	NVRFDIANGS 170	GQTANL 180
		190	200	210	220	230	240
25	orfllng-1.pep	SADYRIVRDHSEPE	EGOGYFTHSYV	GPVVYTPEGN	IFQKVSFSDLD	DDAKSGKSEA 	EYIRKT
	orfl1-1	SADYRIVRDHSEPE	EGQGYFTHSYV	GPVVYTPEGN 210	FOKVSFSDLD 220	DDAKSGKSEA 230	EYIRKT 240
		190	200				
30	orfllng-1.pep	250 PTGWLGMIEHHFMS	260 STWILQPKGG(	270 ONVCAQGDCRI	280 DIKRRNDKLY	290 SASVSVPLTA	300 AIPTRGP
	orf11-1		 STWILQPKGR(	:     :   SVCAAGECN	:	1:11111:1	
35		250	260	270			260
33	orfllng-1.pep	310 KPKMAVNLYAGPO	320 TTSVIANIADI	330 330	340 VHWFASPLFWI	350 LNQLHNIIGN	360 WGWAIV
	orf11-1	: ::       KAEASINLYAGPQ	 TTSVIANIADI	NLQLAKDYGK	VHWFASPLFWI	LNQLHNIIG	NWGWAII
40		300 310	320	330	340	350	
	£11mg-1 non	370 VLTIIVKAVLYPL	380 TNASYRSMAKI	390 MRAAAPKLOT	400 IKEKYGDDRMA	410 AQQQAMMQLYI	420 KDEKINP
	orfllng-1.pep	VLTIIVKAVLYPL	1111111111	1111111111:	1111111111	11111111	111111
45	orf11-1	360 370	380	390	400	410	
		430 LGGCLPMLLQIPV	440	450	460	470	480 ATMEAOT
50	orfllng-l.pep	1111111111111	11111111111	1111111111		11111111	111111
	orf11-1	LGGCLPMLLQIPV 420 430	FIGLYWALFA 440	SVELRQAPWL 450	GWITDLSRAD: 460	470	ATMFAQT
		490	500	510	520	530	540 STEKODA
55	orfllng-1.per			11111111111	1111111111	!!!!!!!!!!	11111
	orf11-1	YLNPPPTDPMQAF 480 490	MMKIMPLVFS 500	VMFFFFPAGI 510	VLYWVVNNLL 520	TIAQQWHINR 530	SIEKQRA
60							
	orfllng-1.pe	p QGEVVSX					
	orfll-1	QGEVVSX 540					
			. C 4 1	_1	on inner me	ombrone pr	otein from

In addition, ORF11ng-1 shows significant homology with an inner-membrane protein from the 65 database (accession number p25754):

5	DT 01-MAY-19 DT 01-NOV-19	U STANDARD; 92 (REL. 22, CRE 92 (REL. 22, LAS 95 (REL. 32, LAS ER-MEMBRANE PROT	T SEQUENCE T ANNOTATIO	UPDATE)	
10	SCORES In Smith-Waterman	it1: 1074 Initn score: 1406;	41.5% iden	: 1103 tity in 574 a	<del>-</del>
15	orfllng-1.pep p25754	11:11 ::1: ::	:  :::	: :   YGQAALPTQNTA	30 40PKPVPAPQQAQKQ        ::: :: ASTVAPGLPDGVPAGNNGASAD 40 50 60
20	orfllng-1.pep	:::::::::::::::::::::::::::::::::::::::	ITVT ::	:::::   : TDVLELAIDPVG	80 90 GDLRRLTLLKYKATGDE-NKPF   :: :     :     GDIVQLNLPKYPRRQDHPNIPF 00 110 120
25	orf11ng-1.pep	VLFGDGKEYTYVAQS	110 1 ELLDAQGNNIL	20 KGIGFSAPK ::   ::  : RASGRPLYAAEQ	130 140 KOYTL-NGDTVEVRLSAPE
30	orfllng-1.pep	150 160 TNGLKIDKVYTFTKD	170 SYLVNVRFDIA   :   :  EYDLNVSYLID	180 NGSGQTANLSAD      :   : NQSGQAWNGNMF	190 200 YRIVRDHS-EPEGQGYF-THSY ::      :  ::   :  AQLKRDASGDPSSSTATGTATY
35	orfllng-1.pep	210 2 VGPVVYTPEGNFQKV	200 20 23 SFSDLDDDAKS !::!:  !:	GKSEAEYIRKTP	220 230  250 260  IGWLGMIEHHFMSTWILOPKGG    :::: : :::    :
40	p25754	LGAALWTASEPYKKV 240 250 270 2	SMKDIDKG 26 80 29	SLKENVS	GGWVAWLQHYFVTAWI-PAKSD 270 280 310 320
45	orfllng-1.pep p25754	:11 :: ::	::   ::	: :: :     GPVISVPA-GGK	PKMAVNLYAGPQTTSVIANIAD : ::      :   : ::: VETSALLYAGPKIQSKLKELSP 320 330
50	orf11ng-1.pep	NLQLAKDYGKVHWF-	ASPLFWLLNQL	1:::11111:1:	370 380 VLTIIVKAVLYPLTNASYRSMA    ::: ::  :       VLTMLIKGLFFPLSAASYRSMA
		350 390	360 400 4	370 10 420	380 390 430 440 LGGCLPMLLQIPVFIGLYWALF
55	p25754	:::::::::::::::::::::::::::::::::::::::	:1111: ::11	11:111 11111	
60		ASVELRQAPWLGWIT	DLSRADPYYIL	1111:1111	490 500 YLNPPPTDPMQAKMMKIMPLVF             :  :
65	p25754	460 470 510	480 520 5	490 30 540	RLNPTPPDPMQAKVMKMMPIIF 500 510
	orfllng-1.pep p25754	SVMFFFFPAGLVLYW ::::::          TFFFLWFPAGLVLYW 520 530	1111 1:1:11	1:1:1 11	

Based on this analysis, including the homology to an inner-membrane protein from *P. putida* and the predicted transmembrane domains (seen in both the meningococcal and gonoccal proteins), it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### 5 Example 8

The following partial DNA sequence was identified in N.meningitidis <SEQ ID 59>:

```
1 ..GCCGTCTTAA TCATCGAATT ATTGACGGGA ACGGTTATC TTTTGGTTGT

51 NAGCGCGGCT TTGGCGGGTT CGGGCATTGC TTACCGGCTG ACCGGCAGTA

101 CGCCTGCCGC CGTCTTGACC GNCGCTCTGC TTTCCGCGCT GGGTATTTNG

151 TTCGTACACG CCAAAACCGC CGTTAGAAAA GTTGAAACGG ATTCATATCA

201 GGATTTGGAT GCCGGACAAT ATGTCGAAAT CCTCCGNCAC ACAGGCGGCA

251 ACCGTTACGA AGTT.TTAT CGCGGTACG. ACTGGCAGGC TCAAAATACG

301 GGGCAAGAAG AGCTTGAACC AGGAACTCGC GCCCTCATTG TCCGCAAGGA

351 AGGCAACCTT CTTATTATCA CACACCCTTA A
```

15 This corresponds to the amino acid sequence <SEQ ID 60; ORF13>:

```
1 ..AVLIIELLTG TVYLLVVSAA LAGSGIAYGL TGSTPAAVLT XALLSALGIX
51 FVHAKTAVRK VETDSYQDLD AGQYVEILRH TGGNRYEVXY RGTXWQAQNT
101 GQEELEPGTR ALIVRKEGNL LIITHP*
```

Further sequence analysis elaborated the DNA sequence slightly <SEQ ID 61>:

	7 (77 477 1	•				
20	51 101 151 201	GCCGTCTTAA nAGCGCGGCT CGCCTGCGCG TTCGTACACG GGATTTGGAT ACCGTTACGA	TTGGCGGGTT CGTCTTGACC CCAAAACCGC GCCGGACAAT	CGGGCATTGC GnCGCTCTGC CGTTAGAAAA ATGTCGAAAT	TTACGGGC1G TTTCCGCGCT GTTGAAACGG CCTCCGACAC ACTGGCAGGC	GGGTATTTNG ATTCATATCA ACAGGCGGCA TCAAAATACG
25	251 301 351	ACCGTTACGA GGGCAAGAAG AGGCAACCTT	AGCTTGAACC	AGGAACTCGC	GCCCTCATTG	TCCGCAAGGA

This corresponds to the amino acid sequence <SEQ ID 62; ORF13-1>:

```
30 LAVLIIELLTG TVYLLVVSAA LAGSGIAYGL TGSTPAAVLT XALLSALGIX
FVHAKTAVRK VETDSYQDLD AGQYVEILRH TGGNRYEVFY RGTHWQAQNT
GQEELEPGTR ALIVRKEGNL LIITHP*
```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from N. meningitidis (strain A)

ORF13 shows 92.9% identity over a 126aa overlap with an ORF (ORF13a) from strain A of N.

35 meningitidis:

55						
		10	20	30	40	50
	orf13.pep	1111111111	YLLVVSAALAGS	11111111111		11111
	orf13a		YLLVVSAALAGS 30	40	50	60
40		10 20	30			
		60 70	80	90	100	110 FRETER
	orf13.pep	VHAKTAVRKVETDSYQDLDAG			111111	, , , , , , ,
45	orf13a	VHAKTAVGKVETDSYQDLDA 70 80	GQYAEILRHAGGN 90	100	110	120
		120				
50	orf13.pep	LIVRKEGNLLIITHPX				

35

```
orf13a LIVRKEGNLLIIAKPX
130
```

The complete length (	DRF13a nucleotide sec	quence <seq 63="" id=""> is:</seq>
-----------------------	-----------------------	------------------------------------

```
ATGACTGTAT GGTTTGTTGC CGCTGTTGCC GTCTTAATCA TCGAATTATT
                     GACGGGAACG GTTTATCTTT TGGTTGTCAG CGCGGCTTTG GCGGGTTCGG
5
                 51
                101
                     GCATTGCTTA CGGGCTGACC GGCAGCACGC CTGCCGCCGT CTTGACCGCC
                     GCTCTGCTTT CCGCGCTGGG TATTTGGTTC GTACACGCCA AAACCGCCGT
                151
               . 201
                     GGGAAAAGTT GAAACGGATT CATATCAGGA TTTGGATGCC GGGCAATATG
                     CCGAAATCCT CCGGCACGCA GGCGGCAACC GTTACGAAGT TTTTTATCGC
                251
10
                     GGTACGCACT GGCAGGCTCA AAATACGGGG CAAGAAGAGC TTGAACCAGG
                301
                     AACGCGCGCC CTAATCGTCC GCAAGGAAGG CAACCTTCTT ATCATCGCAA
                351
                401
                     AACCTTAA
```

### This encodes a protein having amino acid sequence <SEQ ID 64>:

1 MTVWFVAAVA VLIIELLTGT VYLLVVSAAL AGSGIAYGLT GSTPAAVLTA
15 51 ALLSALGIWF VHAKTAVGKV ETDSYQDLDA GQYAEILRHA GGNRYEVFYR
101 GTHWQAQNTG QEELEPGTRA LIVRKEGNLL IIAKP\*

#### ORF13a and ORF13-1 show 94.4% identity in 126 aa overlap

•	orf13a.pep	10 MTVWFVAAVAVLIII	20	30	40	50	60 SALCIME
20 .	Ortioa.beb				111111111		
	orf13-1	AVLIII	ELLTGTVYL:	LVVSAALAGSO	IAYGLTGST	PAAVLTXALL:	SALGIXF
			10	20	30	40	50
	* *						
	•	70	80	90	100	110	120
25	orf13a.pep	VHAKTAVGKVETDS	YQDLDAGQY	<b>AEILRHAGGNE</b>	XEVFYRGTH	WQAQNTGQEE!	LEPGTRA
			1111111	:11111:111	111111111	1111111111	111111
	orf13-1	VHAKTAVRKVETDS	YQDLDAGQY	VEILRHTGGNE	YEVFYRGTH	WQAQNTGQEE!	LEPGTRA
	•	. 60	70	80	90	100	110
30		130					
<b>J</b>	orfl3a.pep	LIVRKEGNLLIIAK	PΧ				
	022200.757	111111111111::	11		•		
	orf13-1	LIVRKEGNLLIITH					
		120					

### Homology with a predicted ORF from N.gonorrhoeae

ORF13 shows 89.7% identity over a 126aa overlap with a predicted ORF (ORF13.ng) from N. gonorrhoeae:

40	orf13	AVLIIELLTGTVYLLVVSAALAGSGIAYGLTGSTPAAVLTXALLSALGIXF	51
40	orf13ng	MTVWFVAAVAVLIIELLTGTVYLLVVSAALAGSGIAYGLTGSTPAAVLTAALLSALGIWF	60
	orf13	VHAKTAVRKVETDSYQDLDAGQYVEILRHTGGNRYEVXYRGTXWQAQNTGQEELEPGTRA	111
45	orf13ng	VHAKTAVGKVETDSYQDLDTGKYAEILRYTGGNRYEVFYRGTHWQAQNTGQEVFEPGTRA	120
	orf13	LIVRKEGNLLIITHP 126	
	orf13ng	LIVRKEGNLLIIANP 135	

#### 50 The complete length ORF13ng nucleotide sequence <SEQ ID 65> is:

	1	ATGACTGTAT	GGTTTGTTGC	CGCTGTTGCC	GTCTTAATCA	TCGAATTATT
	51	GACGGGAACG	GTTTATCTTT	TGGTTGTCAG	CGCGGCTTTG	GCGGGTTCGG
	101	GCATTGCCTA	CGGGCTGACT	GGCAGCACGC	CTGCCGCCGT	CTTGACCGCC
•	<b>1</b> 51	GCACTGCTTT	CCGCGCTGGG	CATTTGGTTC	GTACATGCCA	AAACCGCCGT
55	201	GGGAAAAGTT	GAAACGGATT	CATATCAGGA	TTTGGATACC	GGAAAATATG
	251	CCGAAATCCT	CCGATACACA	GGCGGCAACC	GTTACGAAGT	TTTTTATCGC
	301	GGTACGCACT	GGCAGGCGCA	AAATACGGGG	CAGGAAGTGT	TTGAACCGGG
	. 351	AACGCGCGCC	CTCATCGTCC	GCAAAGAAGG	TAACCTTCTT	ATCATCGCAA
	401	<b>አ</b> ርርርጥጥል ል				

This encodes a protein having amino acid sequence <SEQ ID 66>:

- 1 MTVWFVAAVA VLIIELLTGT VYLLVVSAAL AGSGIAYGLT GSTPAAVLTA 51 ALLSALGIWF VHAKTAVGKV ETDSYQDLDT GKYAEILRYT GGNRYEVFYR
- 101 GTHWQAQNTG QEVFEPGTRA LIVRKEGNLL IIANP\*
- 5 ORF13ng shows 91.3% identity in 126 aa overlap with ORF13-1:

•	 •						
			10	20	30	40	50
	orf13-1.pep	1111111	11111	LLVVSAALAGSG	31111111	11111 1111	1111
10	orf13ng	MTVWFVAAVAVLIIELI 10	TGTVY 20	LLVVSAALAGSG 30	IAYGLTGST 40	PAAVLTAALLS 50	60
10		60	70	80	90	100	110
	orfl3-1.pep	VHAKTAVRKVETDSYQI	111.19	.   •		11111111	
15	orf13ng	VHAKTAVGKVETDSYQI 70	B0	CYAEILRYTGGNI 90	100	110	120
		120					
20	orf13-1.pep	LIVRKEGNLLIITHPX					
20	orf13ng	LIVRKEGNLLIIANPX 130					1

Based on this analysis, including the extensive leader sequence in this protein, it is predicted that ORF13 and ORF13ng are likely to be outer membrane proteins. It is thus predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 9

25

The following DNA sequence was identified in N.meningitidis <SEQ ID 67>:

```
ATGTWTGATT TCGGTTTTGG CGArCTGGTT TTTGTCGGCA TTATCGCCCT
30
                     GATWGECCTC GGCCCCGAAC GCSTGCCCGA GGCCGCCCGC AYCGCCGGAC
                 51
                     GGCTCATCGG CAGGCTGCAA CGCTTTGTCG GCAGCGTCAA ACAGGAATTT
                101
                     GACACTCAAA TCGAACTGGA AGAACTGAGG AAGGCAAAGC AGGAATTTGA
                151
                    AGCTGCCGcC GCTCAGGTTC GAGACAGCCT CAAAGAAACC GGTACGGATA
                     TGGAAGGCAA TCTGCACGAC ATTTCCGACG GTCTGAAGCC TTGGGAAAAA
                201
35
                251
                     CTGCCCGAAC AGCGGACACC TGCCGATTTC GGTGTCGATG AAAACGGCAA
                301
                     TCCGCT.TCC CGATGCGGCA AACACCCTAT CAGACGGCAT TTCCGACGTT
                     ATGCCGTC..
                401
```

This corresponds to the amino acid sequence <SEQ ID 68; ORF2>:

```
40 1 MXDFGLGELV FVGIIALIVL GPERXPEAAR XAGRLIGRLQ RFVGSVKQEF
51 DTQIELEELR KAKQEFEAAA AQVRDSLKET GTDMEGNLHD ISDGLKPWEK
101 LPEQRTPADF GVDENGNPXS RCGKHPIRRH FRRYAV..
```

Further work revealed the complete nucleotide sequence <SEQ ID 69>:

1	51 GATTGTCCTC	TCGGTTTGGG GGCCCCGAAC CAGGCTGCAA TCGAACTGGA	GCCTGCCCGA	GGCCGCCCGC	ACCGCCGGAC
50 3 3 4	AGCTGCCGCC TGGAAGGCAA TCGCCCGAAC TCGCTTCCC TGCCGTCCGA	GCTCAGGTTC TCTGCACGAC AGCGGACACC GATGCGGCAA ACGTTCCTAC GTACAGCCGA	GAGACAGCCT ATTTCCGACG TGCCGATTTC ACACCCTATC GCTTCCGCCG ACCCGCGGAA	CAAAGAAACC GTCTGAAGCC GGTGTCGATG AGACGGCATT AAACCCTTGG ACCGACCAAG	TTGGGAAAAA AAAACGGCAA

	551 601 651	AAGTCAGCTA TATCGATACT GCTGTTGAAA CGCCTGTTCC GCACACCACT TCCCTGCGCA AACAGGCAAT AAGCCGCAAA CGCGATTTTC GTCCGAAACA CCGCGCCAAA CCTAAATTGC GCGTCCGTAA ATCATAA	
	This correspond	ls to the amino acid sequence <seq 70;="" id="" orf2-1="">:</seq>	÷
5	1 51 101 151 201	QTGSTAEPAE TDQDRAWREY LTASAAAPVV QTVEVSYIDT AVETPVPHTT SLRKQAISRK RDFRPKHRAK PKLRVRKS*	
10	Further work id	lentified the corresponding gene in strain A of N. meningitidis < SEQ II	<b>371&gt;:</b>
15	1 51 101 151 201	GACACGCAAA TCGAACTGGA AGAACTAAGG AAGGCAAAGC AGGAATTTGA	
15	251 301 351 401	TGGAGGGTAA TCTGCACGAC ATTTCCGACG GTCTGAAGCC TTGGGAAAAA CTGCCCGAAC AGCGCACGCC TGCTGATTTC GGTGTCGATG AAAACGGCAA TCCCTTTCCC GATGCGGCAA ACACCCTATT AGACGGCATT TCCGACGTTA TGCCGTCCGA ACGTTCCTAC GCTTCCGCCG AAACCCTTGG GGACAGCGGG	
20	451 501 551 601	GCGGGAATAC CTGACTGCTT CTGCCGCCGC ACCCGTCGTA CAGACCGTCG AAGTCAGCTA TATCGATACC GCTGTTGAAA CCCCTGTTCC GCATACCACT	•
	651	CCGCGCCAAA CCTAAATTGC GCGTCCGTAA ATCATAA	
25	This encodes a	protein having amino acid sequence <seq 72;="" id="" orf2a="">:</seq>	• • •
30	1 51 101 151 201	DTQIELEELR KAKQEFEAAA AQVRDSLKET GTDMEGNLHD ISDGLKPWEK LPEQRTPADF GVDENGNPFP DAANTLLDGI SDVMPSERSY ASAETLGDSG QTGSTAEPAE TDQDRAWREY LTASAAAPVV QTVEVSYIDT AVETPVPHTT	
30		-identified partial strain B sequence (ORF2) shows 97.5% identity o	ver a 118aa
30		-identified partial strain B sequence (ORF2) shows 97.5% identity o	ver a 118aa
35	The originally-	identified partial strain B sequence (ORF2) shows 97.5% identity of PRF2a:  10 20 30 40 50	60 ELR
35	The originally- overlap with O  orf2.pep orf2a	identified partial strain B sequence (ORF2) shows 97.5% identity of PRF2a:  10 20 30 40 50  MXDFGLGELVFVGIIALIVLGPERXPEAARXAGRLIGRLQRFVGSVKQEFDTQIELE	60 EELR !!! EELR 60
	The originally- overlap with O	identified partial strain B sequence (ORF2) shows 97.5% identity of PRF2a:  10 20 30 40 50  MXDFGLGELVFVGIIALIVLGPERXPEAARXAGRLIGRLQRFVGSVKQEFDTQIELE	60 EELR !!! EELR 60 120 IPXS
35	The originally- overlap with O  orf2.pep  orf2a	identified partial strain B sequence (ORF2) shows 97.5% identity of the partial strain B seque	60 EELR !!! EELR 60 120 IPXS !!
35 40	The originally- overlap with O  orf2.pep orf2a  orf2.pep orf2a	identified partial strain B sequence (ORF2) shows 97.5% identity of the partial strain B seque	60 EELR !!! EELR 60 120 JPXS !! JPFP 120
35 40	The originally- overlap with O  orf2.pep orf2a  orf2.pep orf2a  orf2.pep orf2a	identified partial strain B sequence (ORF2) shows 97.5% identity of the control o	60 EELR     EELR 60 120 UPXS    NPFP 120
35 40 45	The originally- overlap with O  orf2.pep orf2a  orf2.pep orf2a  orf2.pep orf2a	identified partial strain B sequence (ORF2) shows 97.5% identity of the control o	60 EELR III EELR 60 120 IPXS II IPFP 120 APVV 180 na overlap:
35 40 45 50	The originally- overlap with Originally- overlap with Originally- orf2.pep orf2a  orf2.pep orf2a  orf2.pep orf2a  The complete sorf2a.pep orf2a.pep	identified partial strain B sequence (ORF2) shows 97.5% identity of the control o	APVV 180 APVV 180 APVL 180 APV
35 40 45	The originally- overlap with Originally- overlap with Originally- orf2.pep orf2a  orf2.pep orf2a  orf2.pep orf2a  The complete sorf2a.pep	identified partial strain B sequence (ORF2) shows 97.5% identity of the control o	60 EELR      EELR      EELR     60  120 IPXS     IPFP 120  APVV 180  AB OVERlap: ELEELR 60        ELEELR 60

		-96-
	orf2-1	DAANTLSDGISDVMPSERSYASAETLGDSGQTGSTAEPAETDQDRAWREYLTASAAAPVV 180
	orf2a.pep	QTVEVSYIDTAVETPVPHTTSLRKQAISRKRDLRPKSRAKPKLRVRKSX 229
5	orf2-1	QTVEVSYIDTAVETPVPHTTSLRKQAISRKRDFRPKHRAKPKLRVRKSX 229
	Further work identif	ied a partial DNA sequence <seq 73="" id=""> in N.gonorrhoeae encoding the</seq>
	following amino acid	d sequence <seq 74;="" id="" orf2ng="">:</seq>
10	51 DTO	FGLGELI FVGIIALIVL GPERLPEAAR TAGRLIGRLQ RFVGSVKQEL TELEELR KVKQAFEAAA AQVRDSLKET DTDMQNSLHD ISDGLKPWEK QRTPADF GVDEKGNSLS RYGKHRIRRH FRRYAV*

Further work identified the complete gonococcal gene sequence <SEQ ID 75>:

```
ATGTTTGATT TCGGTTTGGG CGAGCTGATT TTTGTCGGCA TTATCGCCCT
                    GATTGTCCTT GGTCCAGAAC GCCTGCCCGA AGCCGCCCGC ACTGCCGGAC
                51
                    GGCTTATCGG CAGGCTGCAA CGCTTTGTAG GAAGCGTCAA ACAAGAACTT
15
                    GACACTCAAA TCGAACTGGA AGAGCTGAGG AAGGTCAAGC AGGCATTCGA
                151
                    AGCTGCCGCC GCTCAGGTTC GAGACAGCCT CAAAGAAACC GATACGGATA
                201
                    TGCAGAACAG TCTGCACGAC ATTTCCGACG GTCTGAAGCC TTGGGAAAAA
                251
                301 CTGCCCGAAC AGCGCACGCc tgccgatttc gGTGTCGATg AAAacggcaa
                     tececttece gatacggcaa acaccgtate agacggcatt tecgacgtta
                351
20
                    TGCCGTCTGA ACGTTCCGAT ACTTCCGCCG AAACCCTTGG GGACGACAGG
                401
                451 CAAACCGGCA GTACAGCCGA ACCTGCGGAA ACCGACAAAG ACCGCGCATG
                     GCGGGAATAC CTGactgctt ctgccgccgc acctgtcgta Cagagggccg
                501
                     tegaagteag ctaTATCGAT ACTGCTGTTG AAacgcctgT tecgcaCacc
                551
                     acttecetge gcaAACAGGC AATAAACCGC AAACGCGATT TttgtccgaA
25
                601
                     ACACCGCGCc aAACCGAAat tgcgcgtcCG TAAATCATAA
```

This encodes a protein having the amino acid sequence <SEQ ID 76; ORF2ng-1>:

```
1 MFDFGLGELI FVGIIALIVL GPERLPEAAR TAGRLIGRLQ RFVGSVKQEL
                    DTQIELEELR KVKQAFEAAA AQVRDSLKET DTDMQNSLHD ISDGLKPWEK
                    LPEORTPADF GVDENGNPLP DTANTVSDGI SDVMPSERSD TSAETLGDDR
30
               101
                    QTGSTAEPAE TDKDRAWREY LTASAAAPVV QRAVEVSYID TAVETPVPHT
                151
                201 TSLRKQAINR KRDFCPKHRA KPKLRVRKS*
```

The originally-identified partial strain B sequence (ORF2) shows 87.5% identity over a 136aa overlap with ORF2ng:

```
MXDFGLGELVFVGIIALIVLGPERXPEAARXAGRLIGRLQRFVGSVKQEFDTQIELEELR
                                                                         60
35
         orf2.pep
                    i inno: mannon am: manamani: minon
                    MFDFGLGELIFVGIIALIVLGPERLPEAARTAGRLIGRLQRFVGSVKQELDTQIELEELR
         orf2ng
                    KAKQEFEAAAAQVRDSLKETGTDMEGNLHDISDGLKPWEKLPEQRTPADFGVDENGNPXS
         orf2.pep
                    40
                    KVKQAFEAAAAQVRDSLKETDTDMQNSLHDISDGLKPWEKLPEQRTPADFGVDEKGNSLP
         orf2ng
                    RCGKHPIRRHFRRYAV 136
         orf2.pep
                    1 111 1111111111
                    RYGKHRIRRHFRRYAV
45
         orf2ng
```

The complete strain B and gonococcal sequences (ORF2-1 & ORF2ng-1) show 91.7% identity in 229 aa overlap:

	orf2-1.pep	10 MFDFGLGELVFVGII	20 ALIVLGPE	30 RLPEAARTAGR 	40 LIGRLQRFVO	50 SVKQEFDTQI	60 ELEELR
50	orf2ng-1		ALIVLGPE 20	RLPEAARTAGR 30	LIGRLORFVO	SSVKQELDTQI 50	ELEELR 60
55	orf2-1.pep	70 KAKQEFEAAAAQVRD  :            KVKQAFEAAAAQVRD	11111 11	1:::1111111	111111111	11111111111	11111

15

30

35

		70	80	90	100	110	120
		130	140	150	160	170	180
_	orf2-1.pep	DAANTLSDGISDVN		-	_		
5		1:111:1111111			11111111:1		
	orf2ng-1	DTANTVSDGISDV	1PSERSDTSAE	TLGDDRQTGS	TAEPAETDKE	RAWREYLTA	SAAAPVV
		130	140	150	160	170	180
		190	200	210	220	229	•
10	orf2-1.pep	Q-TVEVSYIDTAVE	ETPVPHTTSLE	KQAISRKRDE	RPKHRAKPKI	RVRKSX	
		1:1111111111		1111:1111	111111111	111111	
	orf2ng-1	QRAVEVSYIDTAVI	ETPVPHTTSLP	KQAINRKRDI	CPKHRAKPKI	RVRKSX	
	_	. 190	200	210	220	230	

Computer analysis of these amino acid sequences indicates a transmembrane region (underlined), and also revealed homology (59% identity) between the gonococcal sequence and the TatB protein of *E.coli*:

```
gnl|PID|e1292181 (AJ005830) TatB protein [Escherichia coli] Length = 171
           Score = 56.6 bits (134), Expect = 1e-07
           Identities = 30/88 (34%), Positives = 52/88 (59%), Gaps = 1/88 (1%)
20
          Query: 1 MFDFGLGELIFVGIIALIVLGPERLPEAARTAGRLIGRLQRFVGSVKQELDTQIELEELR 60
                    MFD G EL+ V II L+VLGP+RLP A +T
                                                       I L+
                                                                +V+ EL +++L+E +
          Sbjct: 1 MFDIGFSELLLVFIIGLVVLGPQRLPVAVKTVAGWIRALRSLATTVQNELTQELKLQEFQ 60
25
          Query: 61 -KVKQAFEAAAAQVRDSLKETDTDMQNS 87
                       +K+
                           +A+
                                 +
                                     LK + +++ +
          Sbjct: 61 DSLKKVEKASLTNLTPELKASMDELRQA 88
```

Based on this analysis, it was predicted that ORF2, ORF2a and ORF2ng are likely to be membrane proteins and so the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF2-1 (16kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 3A shows the results of affinity purification of the GST-fusion protein, and Figure 3B shows the results of expression of the His-fusion in *E.coli*. Purified GST-fusion protein was used to immunise mice, whose sera were used for Western blots (Figure 3C), ELISA (positive result), and FACS analysis (Figure 3D). These experiments confirm that ORF37-1 is a surface-exposed protein, and that it is a useful immunogen.

#### Example 10

The following partial DNA sequence was identified in N. meningitidis <SEO ID 77>:

40	1	ATGCAAGCAC	GGCTGCTGAT	ACCTATTCTT	TTTTCAGTTT	TTATTTTATC
	51				GCATGGCGGA	
	101	TTgCGGTCGA	ACAAGAACTT	GTGGCCGCTT	CTGCCAGAGC	TGCCGTTAAA
	151	GACATGGATT	TACAGGCATT	ACACGGACGA	AAAGTTGCAT	TGTACATTGC
	201	CACTATGGGC	GACCAAGGTT	CAGGCAGTTT	GACAGGGGG	TCGCTACTCC
45	251	ATTGATGCAC	kGrTwCsTGG	CGAATACATA	AACAGCCCTG	CCGTCCGTAC
	301					ACATCAGGCG
	351					CCCTGCACTC
•	401	TCTCGCACCC	AATCAGACGG	TAGCGGAAGT	AAAAGCAGTC	TGGGCTTAAA
	451	TATTGGCGGG	ATGGGGGATT	ATCGAAATGA	AACCTTGACG	ACTAACCCGC

```
GCGACACTGC CTTTCTTTCC CACTTGGTAC AGACCGTATT TTTCCTGCGC
                    GGCATAGACG TTGTTTCTCC TGCCAATGCC GATACAGATG TGTTTATTAA
               551
                    CATCGACGTA TTCGGAACGA TACGCAACAG AACCGAAATG..
               601
     This corresponds to the amino acid sequence <SEQ ID 78; ORF15>:
                 1 MQARLLIPIL FSVFILSACG TLTGIPSHGG XKRFAVEQEL VAASARAAVK
5
                     DMDLQALHGR KVALYIATMG DQGSGSLTGG RYSIDAXXXG EYINSPAVRT
                     DYTYPRYETT AETTSGGLTG LTTSLSTLNA PALSRTQSDG SGSKSSLGLN
               101
                     IGGMGDYRNE TLTTNPRDTA FLSHLVQTVF FLRGIDVVSP ANADTDVFIN
                201 IDVFGTIRNR TEM..
     Further work revealed the complete nucleotide sequence <SEQ ID 79>:
10
                  1 ATGCAAGCAC GGCTGCTGAT ACCTATTCTT TTTTCAGTTT TTATTTTATC
                     CGCCTGCGGG ACACTGACAG GTATTCCATC GCATGGCGGA GGTAAACGCT
                 51
                     TTGCGGTCGA ACAAGAACTT GTGGCCGCTT CTGCCAGAGC TGCCGTTAAA
                101
                     GACATGGATT TACAGGCATT ACACGGACGA AAAGTTGCAT TGTACATTGC
                151
                     CACTATGGGC GACCAAGGTT CAGGCAGTTT GACAGGGGGT CGCTACTCCA
15
                     TTGATGCACT GATTCGTGGC GAATACATAA ACAGCCCTGC CGTCCGTACC
                251
                     GATTACACCT ATCCACGTTA CGAAACCACC GCTGAAACAA CATCAGGCGG
                301
                     TTTGACAGGT TTAACCACTT CTTTATCTAC ACTTAATGCC CCTGCACTCT
                351
                     CTCGCACCCA ATCAGACGGT AGCGGAAGTA AAAGCAGTCT GGGCTTAAAT
                 401
                451 ATTGGCGGGA TGGGGGATTA TCGAAATGAA ACCTTGACGA CTAACCCGCG
20
                      CGACACTGCC TTTCTTTCCC ACTTGGTACA GACCGTATTT TTCCTGCGCG
                 501
                      GCATAGACGT TGTTTCTCCT GCCAATGCCG ATACAGATGT GTTTATTAAC
                 551
                      ATCGACGTAT TCGGAACGAT ACGCAACAGA ACCGAAATGC ACCTATACAA
                 601
                      TGCCGAAACA CTGAAAGCCC AAACAAAACT GGAATATTTC GCAGTAGACA
                 651
                      GAACCAATAA AAAATTGCTC ATCAAACCAA AAACCAATGC GTTTGAAGCT
25
                 701
                      GCCTATAAAG AAAATTACGC ATTGTGGATG GGGCCGTATA AAGTAAGCAA
                 751
                      AGGAATTAAA CCGACGGAAG GATTAATGGT CGATTTCTCC GATATCCGAC
                 801
                      CATACGGCAA TCATACGGGT AACTCCGCCC CATCCGTAGA GGCTGATAAC
                 851
                      AGTCATGAGG GGTATGGATA CAGCGATGAA GTAGTGCGAC AACATAGACA
                 901
                      AGGACAACCT TGA
 30
                 951
      This corresponds to the amino acid sequence <SEQ ID 80; ORF15-1>:
                      MOARLLIPIL FSVFILSACG TLTGIPSHGG GKRFAVEQEL VAASARAAVK
                      DMDLQALHGR KVALYIATMG DQGSGSLTGG RYSIDALIRG EYINSPAVRT
                      DYTYPRYETT AETTSGGLTG LTTSLSTLNA PALSRTQSDG SGSKSSLGLN
                 101
                      IGGMGDYRNE TLTTNPRDTA FLSHLVQTVF FLRGIDVVSP ANADTDVFIN
 35
                 151
                      IDVFGTIRNR TEMHLYNAET LKAQTKLEYF AVDRTNKKLL IKPKTNAFEA
                      AYKENYALWM GPYKVSKGIK PTEGLMVDFS DIRPYGNHTG NSAPSVEADN
                 251
                  301 SHEGYGYSDE VVRQHRQGQP *
       Further work identified the corresponding gene in strain A of N. meningitidis <SEQ ID 81>:
                   1 ATGCAAGCAC GGCTGCTGAT ACCTATTCTT TTTTCAGTTT TTATTTTATC
 40
                      CGCCTGCGGG ACACTGACAG GTATTCCATC GCATGGCGGA GGTAAACGCT
                      TTGCGGTCGA ACAAGAACTT GTGGCCGCTT CTGCCAGAGC TGCCGTTAAA
                  101
                       GACATGGATT TACAGGCATT ACACGGACGA AAAGTTGCAT TGTACATTGC
                      AACTATGGGC GACCAAGGTT CAGGCAGTTT GACAGGGGGT CGCTACTCCA
                  201
                       TTGATGCACT GATTCGTGGC GAATACATAA ACAGCCCTGC CGTCCGTACC
  45
                      GATTACACCT ATCCACGTTA CGAAACCACC GCTGAAACAA CATCAGGCGG
                  301
                       TTTGACAGGT TTAACCACTT CTTTATCTAC ACTTAATGCC CCTGCACTCT
                  351
                       CGCGCACCCA ATCAGACGGT AGCGGAAGTA AAAGCAGTCT GGGCTTAAAT
                  401
                       ATTGGCGGGA TGGGGGATTA TCGAAATGAA ACCTTGACGA CTAACCCGCG
                  451
                       CGACACTGCC TTTCTTTCCC ACTTGGTACA GACCGTATTT TTCCTGCGCG
  50
                       GCATAGACGT TGTTTCTCCT GCCAATGCCG ATACGGATGT GTTTATTAAC ATCGACGTAT TCGGAACGAT ACGCAACAGA ACCGAAATGC ACCTATACAA
                  551
                  601
                       TGCCGAAACA CTGAAAGCCC AAACAAAACT GGAATATTTC GCAGTAGACA
                  651
                       GAACCAATAA AAAATTGCTC ATCAAACCAA AAACCAATGC GTTTGAAGCT
                  701
                       GCCTATAAAG AAAATTACGC ATTGTGGATG GGACCGTATA AAGTAAGCAA
                  751
  55
                       AGGAATTAAA CCGACAGAAG GATTAATGGT CGATTTCTCC GATATCCAAC
                  801
                       CATACGGCAA TCATATGGGT AACTCTGCCC CATCCGTAGA GGCTGATAAC
                  851
```

AGGGCAACCT TGA

AGTCATGAGG GGTATGGATA CAGCGATGAA GCAGTGCGAC GACATAGACA

<sup>60</sup> This encodes a protein having amino acid sequence <SEQ ID 82; ORF15a>:

<sup>1</sup> MOARLLIPIL FSVFILSACG TLTGIPSHGG GKRFAVEQEL VAASARAAVK

51 DMDLQALHGR KVALYIATMG DQGSGSLTGG RYSIDALIRG EYINSPAVRT
101 DYTYPRYETT AETTSGGLTG LTTSLSTLNA PALSRTQSDG SGSKSSLGLN
151 IGGMGDYRNE TLTTNPRDTA FLSHLVQTVF FLRGIDVVSP ANADTDVFIN
201 IDVFGTIRNR TEMHLYNAET LKAQTKLEYF AVDRTNKKLL IKPKTNAFEA
5 251 AYKENYALWM GPYKVSKGIK PTEGLMVDFS DIQPYGNHMG NSAPSVEADN
301 SHEGYGYSDE AVRRHRQGQP \*

The originally-identified partial strain B sequence (ORF15) shows 98.1% identity over a 213aa overlap with ORF15a:

		10	. 20	30	40	50	60
10	orf15.pep	MOARLLIPILFSVF	ILSACGTLT	GIPSHGGXKRF	AVEQELVAAS	SARAAVKDMDL	QALHGR
	-		Ш	11111111111	1111111111		
	orf15a	MQARLLIPILFSVF	ILSACGTLT	GIPSHGGGKRF	AVEQELVAAS	SARAAVKDMDL	QALHGR
	•	10	20	30	40	50	60
15		70	. 80	90	100	110	120
	orf15.pep	KVALYIATMGDQGS	GSLTGGRYS	IDAXXXGEYIN	SPAVRTDYTY	PRYETTAETT	SGGLTG
	_		HITTHEE	111 11111	11111111111		111111
	orf15a	KVALYIATMGDQGS	GSLTGGRYS	IDALIRGEYIN	SPAVRTDYTY	PRYETTAETT	SGGLTG
		70	80	90	100	110	120
20							
	• •	130	140	150	160	170	180
	orf15.pep	LTTSLSTLNAPALS	RTQSDGSGS	KSSLGLNIGGM	GDYRNETLTI	INPRDTAFLSH	LVQTVF
			111111111	1111111111	$\mathbf{H}\mathbf{H}\mathbf{H}\mathbf{H}\mathbf{H}$	11111111111	HIIII
•	orf15a	LTTSLSTLNAPALS		KSSLGLNIGGM	GDYRNETLT	Inprdtaflsh	LVQTVF
25		· 130	140	150	160	170	180
	٠.						
	•	190	200	210			
	orf15.pep	FLRGIDVVSPANAD	TDVFINIDV	FGTIRNRTEM			
			11111111	11111111			
30	orf15a	FLRGIDVVSPANAD			<del>-</del>		
	•	190	200	210	220	230	240

The complete strain B sequence (ORF15-1) and ORF15a show 98.8% identity in 320 aa overlap:

		10	20	30	40	50	60
25	orf15a.pep	MQARLLIPILFSVFI					
35	orf15-1		· · · · · · · · ·				
	01112-1	10	20	30	40	50	60
		20	20	30	10	30	•
		70	80	90	100	110	120
40	orf15a.pep	KVALYIATMGDQGSG	SLTGGRYSI	DALIRGEYIN	SPAVRTDYTY	PRYETTAETT	SGGLTG
			111111111	111111111	111111111	1111111111	
	orf15-1	KVALYIATMGDQGSG					
		70	80	90	100	110	120
45		130	140	150	160	170	180
	orf15a.pep	LTTSLSTLNAPALSR					
			11111111	3111111111	11111111111	1111111111	HIĪH
	orf15-1	LTTSLSTLNAPALSP					
50		130	140	150	160	170	180
50		190	200	210	220	230	240
	orf15a.pep	FLRGIDVVSPANADI					
	Olliou.pop						
	orf15-1	FLRGIDVVSPANADT			· · · · · · · · · · · ·		
55		190	200	210	220	230	240
	C1 C	250	260	270	280	290	300
	orf15a.pep	IKPKTNAFEAAYKEN	NIALWMGPI	CVSKGIKPTEC	PEWADESDIGE		PSVEADN
60	orf15-1	IKPKTNAFEAAYKEN	IYAT.WMGPYI	CVSKGTKPTEC	IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII		
	01110 1	250	260	270	280	290	300
					•		
·	•	310	320				
	orf15a.pep	SHEGYGYSDEAVRRI				•	
65							•
	orf15-1	SHEGYGYSDEVVRQ	TKÔPŐ P.Y				

PCT/IB98/01665

310 320

```
Further work identified the corresponding gene in N.gonorrhoeae <SEQ ID 83>:
```

	I diction		_			
5	1 51 101 151	ATGCGGGCAC CGCCTGCGGG TCGCGGTCGA GACATGGATT AACTATGGGC	ACACTGACAG ACAAGAACTT TACAGGCATT	GTATTCCATC GTGGCCGCTT ACACGGACGA	AAAGTTGCAT GACAGGGGGT	TGCCGTTAAA TGTACATTGC CGCTACTCCA
10	251 301 351 401 451 501	TTGATGCACT GATTACACCT TTTGACGGGT CGCGCACCCA ATTGGCGGGA	GATTCGCGGC ATCCGCGTTA	CGAAACCACC CTTTATCTAC AGCGGAAGTA TCGAAATGAA ACTTGGTGCA	GCTGAAACAA ACTTAATGCC GGAGCAGTCT ACCTTGACGA GACCGTATTT	CATCAGGCGG CCTGCACTCT GGGCTTAAAT CCAACCCGCG TTCCTGCGCG
15	551 601 651 701 751	GCATAGACGT ATCGACGTAT TGCCGAAACA GAACCAATAA GCCTATAAAG	TGTTTCTCCT TCGGAACGAT CTGAAAGCCC AAAATTGCTC AAAATTACGC	GCCAATGCCG ACGCAACAGA AAACAAAACT		ACCTATACAA GCAGTAGACA
20	801 851 901 951	CATACGGCAA	TCATACGGGT GGTATGGATA	AACTCCGCCC	CATCCGTAGA	GGCTGATAAC AACATAGACA

# This encodes a protein having amino acid sequence <SEQ ID 84; ORF15ng>:

```
25 MRARLLIPIL FSVFILSACG TLTGIPSHGG GKRFAVEQEL VAASARAAVK
51 DMDLQALHGR KVALYIATMG DQGSGSLTGG RYSIDALIRG EYINSPAVRT
101 DYTYPRYETT AETTSGGLTG LTTSLSTLNA PALSRTQSDG SGSRSSLGLN
151 IGGMGDYRNE TLTTNPRDTA FLSHLVQTVF FLRGIDVVSP ANADTDVFIN
201 IDVFGTIRNR TEMHLYNAET LKAQTKLEYF AVDRTNKKLL IKPKTNAFEA
251 AYKENYALWM GPYKVSKGIK PTEGLMVDFS DIQPYGNHTG NSAPSVEADN
30 301 SHEGYGYSDE AVRQHRQGQP *
```

The originally-identified partial strain B sequence (ORF15) shows 97.2% identity over a 213aa overlap with ORF15ng:

25	orf15.pep	MQARLLIPILFSVFILSACGTLTGIPSHGGXKRFAVEQELVAASARAAVKDMDLQALHGR  :	60	
35	orribing			
	orf15.pep	KVALYIATMGDQGSGSLTGGRYSIDAXXXGEYINSPAVRTDYTYPRYETTAETTSGGLTG	120 120	
	orf15ng	KVALYIATMGDQGSGSLTGGRYSIDALIRGEYINSPAVRTDYTYPRYETTAETTSGGLTG	120	
40	orf15.pep	LTTSLSTLNAPALSRTQSDGSGSKSSLGLNIGGMGDYRNETLTTNPRDTAFLSHLVQT		
	orf15ng	LTTSLSTLNAPALSRTQSDGSGSRSSLGLNIGGMGDYRNETLTTNPRDTAFLSHLVQTVF	180	
4.5	£1.E	FLRGIDVVSPANADTDVFINIDVFGTIRNRTEM	213	
45	orf15.pep		240	
			1	

The complete strain B sequence (ORF15-1) and ORF15ng show 98.8% identity in 320 aa overlap:

	-						
		10	20	30	40	50	60
50	orf15-1.pep	MOARLLIPILFSVFI		11111111111	111111111	11111111	11111
	orf15ng	MRARLLIPILFSVF	LSACGTLT 20	GIPSHGGGKRF. 30	AVEQELVAAS 40	ARAAVKDMDL 50	QALHGR 60
		10	20	50			
55		70	80	90	100	110	120
33	orf15-1.pep	KVALYIATMGDQGS	SSLTGGRYS	IDALIRGEYIN	SPAVRTDYT	/PRYETTAETI	SGGLIG
	<b>-4.6</b>			ITDALIRGEYIN	SPAVRTDYT	YPRYETTAETI	SGGLTG
	orf15ng	70	80	90	100	110	120
60		130	140	150	160	170	180
	orf15-1.pep	LTTSLSTLNAPALS	RTQSDGSGS	KSSLGLNIGGN	IGDYRNETLT	TNPKDTAFLSI	Trootar.

	orf15ng	LTTSLS	 TLNAPALSRT 130	   DSDGSGSRSS   140				 
5			190	200	210	220	230	240
_	orf15-1.pep	FLRGID	VVSPANADTD'	VFINIDVFGT	IRNRTEMHLY	NAETĻKAQTK	LEYFAVDRTN	KKLL
		111111	1111111111		111111111	11111111	[]	
	orf15ng	ETERCID	VVSPANADTD 190	200	IRNKTEMHLI 210	NAETLKAUTK. 220	LEYFAVDRTNI 230	240
10	,						230	240
	•		250	260	270	280	290	300
	orf15-1.pep	IKPKTN	AFEAAYKENY.	ALWMGPYKVS	KGIKPTEGLM	VDFSDIRPYG	nhtgnsapsvi	EADN
					11111111111	111:111	1111111111	
	orf15ng	IKPKTN				_	nhtgnsapsvi	EADN
15			250	260	270	280	290	300
	•		310	320			•	
	orf15-1.pep	SHEGYG	YSDEVVRQHR	QGQPX				
		111111	1111:1111	HIH				
20	orf15ng	SHEGYG	YSDEAVROHR	QGQPX				
			310	320				

Computer analysis of these amino acid sequences reveals an ILSAC motif (putative membrane lipoprotein lipid attachment site, as predicted by the MOTIFS program).

indicates a putative leader sequence, and it was predicted that the proteins from N.meningitidis and 25 N.gonorrhoeae, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF15-1 (31.7kDa) was cloned in pET and pGex vectors and expressed in E.coli, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 4A shows the results of affinity purification of the GST-fusion protein, and Figure 4B shows the results of expression of the His-fusion in E.coli. Purified GST-fusion protein was used to immunise mice, whose sera were used for Western blot (Figure 4C) and ELISA (positive result). These experiments confirm that ORFX-1 is a surface-exposed protein, and that it is a useful immunogen.

#### Example 11

30

The following partial DNA sequence was identified in N. meningitidis <SEQ ID 85>:

35 ·	1	GG.CAGCACA	AAAAACAGGC	GGTTGAACGG	AAAAACCGTA	TTTACGATGA
	51	TGCCGGGTAT	GATATTCGGC	GTATTCACGG	GCGCATTCTC	CGCAAAATAT
	101	ATCCCCGCGT	TCGGGCTTCA	AATTTTCTTC	ATCCTGTTTT	TAACCGCCGT
	151	CGCATTCAAA	ACACTGCATA	CCGACCCTCA	GACGGCATCC	CGCCCGCTGC
	201	CCGGACTGCC	CrGACTGACT	GCGGTTTCCA	CACTGTTCGG	CACAATGTCG
40	251	AGCTGGGTCG	GCATAGGCGG	CGGTTCACTT	TCCGTCCCCT	TCTTAATCCA
	301	CTGCGGCTTC	CCCGCCCATA	<b>AAGCCATCGG</b>	CACATCATCC	GGCCTTGCCT
	. 351	GGCCGATTGC	ACTCTCCGGC	GCAATATCGT	ATCTGCTCAA	CGGCCTGAAT
	401	ATTGCAGGAT	TGCCCGAAGG	GTCACTGGGC	TTCCTTTACC	TGCCCGCCGT
	451	CGCCGTCCTC	AGCGCGGCAA	CCATTGCCTT	TGCCCCGCTC	GGTGTCAAAA
45	501	CCGCCCACAA	ACTTTCTTCT	GCCAAACTCA	AAAAATC.TT	CGGCATTATG
	551	TTGCTTTTGA	TTGCCGGAAA	AATGCTGTAC	<b>AACCTGCTTT</b>	AA

This corresponds to the amino acid sequence <SEQ ID 86; ORF17>:

- .GQHKKQAVNG KTVFTMMPGM IFGVFTGAFS AKYIPAFGLQ IFFILFLTAV
- AFKTLHTDPQ TASRPLPGLP XLTAVSTLFG TMSSWVGIGG GSLSVPFLIH

```
101 CGFPAHKAIG TSSGLAWPIA LSGAISYLLN GLNIAGLPEG SLGFLYLPAV
151 AVLSAATIAF APLGVKTAHK LSSAKLKKSF GIMLLLIAGK MLYNLL*
```

Further work revealed the complete nucleotide sequence <SEQ ID 87>:

```
1 ATGTGGCATT GGGACATTAT CTTAATCCTG CTTGCCGTAG GCAGTGCGGC
                    AGGTTTTATT GCCGGCCTGT TCGGCGTAGG CGGCGGCACG CTGATTGTCC
5
                 51
                    CTGTCGTTTT ATGGGTGCTT GATTTGCAGG GTTTGGCACA ACATCCTTAC
                101
                    GCGCAACACC TCGCCGTCGG CACATCCTTC GCCGTCATGG TCTTCACCGC
                151
                    CTTTTCCAGT ATGCTGGGGC AGCACAAAAA ACAGGCGGTC GACTGGAAAA
                201
                    CCGTATTTAC GATGATGCCG GGTATGATAT TCGGCGTATT CACGGGCGCA
                251
                     CTCTCCGCAA AATATATCCC CGCGTTCGGG CTTCAAATTT TCTTCATCCT
                301
10
                    GTTTTTAACC GCCGTCGCAT TCAAAACACT GCATACCGAC CCTCAGACGG
                351
                     CATCCCGCCC GCTGCCCGGA CTGCCCGGAC TGACTGCGGT TTCCACACTG
                401
                     TTCGGCACAA TGTCGAGCTG GGTCGGCATA GGCGGCGGTT CACTTTCCGT
                451
                     CCCCTTCTTA ATCCACTGCG GCTTCCCCGC CCATAAAGCC ATCGGCACAT
                501
                     CATCCGGCCT TGCCTGGCCG ATTGCACTCT CCGGCGCAAT ATCGTATCTG
                551
15
                     CTCAACGGCC TGAATATTGC AGGATTGCCC GAAGGGTCAC TGGGCTTCCT
                601
                     TTACCTGCCC GCCGTCGCCG TCCTCAGCGC GGCAACCATT GCCTTTGCCC
                651
                     CGCTCGGTGT CAAAACCGCC CACAAACTTT CTTCTGCCAA ACTCAAAAAA
                701
                     TC.TTCGGCA TTATGTTGCT TTTGATTGCC GGAAAAATGC TGTACAACCT
                751
                801 GCTTTAA
20
```

This corresponds to the amino acid sequence <SEQ ID 88; ORF17-1>:

```
25 MWHWDIILIL LAVGSAAGFI AGLFGVGGGT LIVPVVLWVL DLQGLAQHPY
51 AQHLAVGTSF AVMVFTAFSS MLGQHKKQAV DWKTVFTMMP GMIFGVFTGA
101 LSAKYIPAFG LQIFFILFLT AVAFKTLHTD PQTASRPLPG LPGLTAVSTL
151 FGTMSSWVGI GGGSLSVPFL IHCGFPAHKA IGTSSGLAWP IALSGAISYL
201 LNGLNIAGLP EGSLGFLYLP AVAVLSAATI AFAPLGVKTA HKLSSAKLKK
251 XFGIMLLLIA GKMLYNLL*
```

Computer analysis of this amino acid sequence gave the following results:

Homology with hypothetical H.influenzae transmembrane protein HI0902 (accession number P44070)

30 ORF17 and HI0902 proteins show 28% aa identity in 192 aa overlap:

```
HKKQAVNGKTVFTMMPGMIFGVFT-GAFSAKYIPAFGLQIF--FILFLTAVAFKTLHTDP 59
          ORF17
                           + + V + P ++ VF G F +
                                                           +IF
                                                                +++L
                     HKLGNIVWQAVRILAPVIMLSVFICGLFIGRLDREISAKIFACLVVYLATKMVLSIKKD- 130
          HI0902
                  60 QTASRPLPGLPXLTAVSTLFGTMSSWVGIGGGSLSVPFLIHCGFPAHKAIGTSSGLAWPI 119
35
          ORF17
                                       L G SS GIGGG VPFL G
                                                                    +AIG+S+
                 131 QVTTKSLTPLSSVIG-GILIGMASSAAGIGGGGFIVPFLTARGINIKQAIGSSAFCGMLL 189
          HI0902
                  120 ALSGAISYLLNGLNIAGLPEGSLGFLYLPAVAVLSAATIAFAPLGVXXXXXXXXXXXXXXX 179
          ORF17
                                      +PE SLG++YLPAV ++A +
                       +SG S++++G
40
                 190 GISGMFSFIVSGWGNPLMPEYSLGYIYLPAVLGITATSFFTSKLGASATAKLPVSTLKKG 249
          HI0902
                  180 FGIMLLLIAGKM 191
          ORF17
                      F + L+++A M
           HI0902 250 FALFLIVVAINM 261
45
```

# Homology with a predicted ORF from N. meningitidis (strain A)

ORF17 shows 96.9% identity over a 196aa overlap with an ORF (ORF17a) from strain A of N. meningitidis:

50					10	20	30
30	orf17.pep			GQHI	KKQAVNGKT <u>V</u>	FTMMPGMIFG	<u>VFTGA</u> FS
	Offi, beb			111	$\Pi\Pi\Pi:\Pi\overline{\Pi}$	1111111:11	11:11:1
	C17-	OGLAOHPYAOHLA	VCTSFAVMVF	PAFSSMLGOH	KKOAVDWKTV	FTMMPGMVFG	VFAGALS
	orf17a	50	60	70	80	90	100
55			50	60	70	80	90
		40	50	60			
	orf17.pep	AKYIPAFGLQIFE	1111111111	111111111	11111111111	1111111111	111111
	orf17a	AKYIPAFGLQIFF	ILFLTAVAFK	TLHTDPQTAS	RPLPGLPGLT	AVSTLFGTMS	SSWVGIGG

-103-

		110	120	130	140	150	160
5	orf17.pep orf17a	100 GSLSVPFLIHCGI           GSLSVPFLIHCGI 170		1111111111	1111111111	1111111111	шинП
10	orf17.pep orf17a	160 AVLSAATIAFAPI IIIIIIIIIIIIIII AVLSAATIAFAPI 230	ПППП	$\Pi\Pi\Pi\Pi\Pi\Pi$	111111111	11111	
	The complete lengt	h ORF17a nucl	eotide sec	quence <se< th=""><th>Q ID 89&gt; i</th><th>s:</th><th>,</th></se<>	Q ID 89> i	s:	,
15	51 AG	GTGGCATT GGGAGGTTTTATT GCCGG	SCCTGT TO	GGCGTAGG C	GGCGGCACG	CTGATTGTCC	
20	151 GC 201 CT 251 CC 301 CT 351 GT	GCAACACC TCGCC TTTCCAGT ATGCT GTATTTAC GATGA CTCCGCAA AATAT TTTTAACC GCCGT	CGTCGG CA CGGGGC AG ATGCCG GG PATCCC AG CGCAT TC	CATCCTTC G CACAAAAA A TATGGTAT T CGTTCGGG C AAAACACT G	CCGTCATGG CAGGCGGTC CGGCGTATT TTCAAATTT CATACCGAC	TCTTCACCGC GACTGGAAAA CGCTGGCGCA TCTTCATCCT CCTCAGACGG	
25	451 TT 501 CC 551 CA 601 CT	TCCCGCCC GCTGC CGGCACAA TGTCC CCTTCTTA ATCCA TCCGGCCT TGCCT CAACGGCC TGAAT ACCTGCCC GCCGT	SAGCTG GG ACTGCG GC IGGCCG AT FATTGC AG	TCGGCATA G TTCCCCGC C TGCACTCT C GATTGCCC G	GCGGCGGTT CATAAAGCC CGGCGCAAT AAGGGTCAC	CACTTTCCGT ATCGGCACAT ATCGTATCTG TGGGCTTCCT	
30	701 CG 751 TC	CTCGGTGT CAAAA CTTCGGCA TTATC	ACCGCC CA	CAAACTTT C	TTCTGCCAA	ACTCAAAAAA	
	This encodes a prot	tein having ami	no acid se	quence <sf< th=""><th>EQ ID 90&gt;:</th><th></th><th></th></sf<>	EQ ID 90>:		
35	51 ĀQ 101 LS 151 FG 201 LN	HWDIILIL LAVGS HLAVGTSF AVMVI AKYIPAFG LQIFI TMSSWVGI GGGSI GLNIAGLP EGSLO GIMLLLIA GKMLY	TTAFSS ML FILFLT AV LSVPFL IH GFLYLP AV	GQHKKQAV D AFKTLHTD P CGFPAHKA I	WKTVFTMMP QTASRPLPG GTSSGLAWP	GMVFGVFAGA LPGLTAVSTL IALSGAISYL	
	ORF17a and ORF1			in 268 aa o	verlap:		
40	orf17a.pep	10 MWHWDIILILI	2 LAVGSAAGF	0 30	0 4 TLIVPVVLWV	0 50 LDLQGLAQHPY	AOHLAVGTSF
4	orf17-1	MWHWDIILILI 10	LAVGSAAGF	IAGLFGVGGG	TLIVPVVLWV	LDLQGLAQHPY 0 50	AQHLAVGTSF
45	orf17a.pep	70 AVMVFTAFSSN	<b>ILGQHKKQA</b>	0 90 VDWKTVFTMM	PGMVFGVFAG	ALSAKYIPAFO	LOIFFILFLT
50	orf17-1	AVMVFTAFSSN 70	<b>1</b> LGQHKKQA	VDWKTVFTMM 0 9	<b>PGMI FGVFT</b> G	HIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	LQIFFILFLT
	orf17a.pep	130 AVAFKTLHTDI	POTASRPLP	0 150 GLPGLTAVST	LFGTMSSWVG	IGGGSLSVPFI	IHCGFPAHKA
55	orf17-1	AVAFKTLHTDI 130	OTASRPLP 14	GLPGLTAVST	LFGTMSSWVG		IHCGFPAHKA
60	orf17a.pep	111111111		LLNGLNIAGL	PEGSLGFLYL	0 230 PAVAVLSAATI            PAVAVLSAATI	AFAPLGVKTA
	OLLIV-1	190	20				
65	orf17a.pep	250 HKLSSAKLKKS	26 FGIMLLLI			•	

-104-

5 Homology with a predicted ORF from N. gonorrhoeae

ORF17 shows 93.9% identity over a 196aa overlap with a predicted ORF (ORF17.ng) from N. gonorrhoeae:

	orf17.pep	GQHKKQAVNGKTVFTMMPGMIFGVFTGAFS	30
10	orf17ng	QGLAQHPYAQHLAVGTSFAVMVFTAFSSMLGQHKKQAVDWKTIFAMMPGMIFGVFAGALS	102
	orf17.pep	AKYIPAFGLQIFFILFLTAVAFKTLHTDPQTASRPLPGLPXLTAVSTLFGTMSSWVGIGG	90
15	orf17ng	AKYIPAFGLQIFFILFLTAVAFKTLHTGRQTASRPLPGLPGLTAVSTLFGAMSSWVGIGG	162
	orf17.pep	GSLSVPFLIHCGFPAHKAIGTSSGLAWPIALSGAISYLLNGLNIAGLPEGSLGFLYLPAV	
	orf17ng	GSLSVPFLIHCGFPAHKAIGTSSGLAWPIALSGAISYLVNGLNIAGLPEGSLGFLYLPAV	202
20	orf17.pep	AVLSAATIAFAPLGVKTAHKLSSAKLKKSFGIMLLLIAGKMLYNLL 196	
	orf17ng	AVLSAATIAFAPLGVKTAHKLSSAKLKESFGIMLLLIAGKMLYNLL 268	

An ORF17ng nucleotide sequence <SEQ ID 91> is predicted to encode a protein having amino acid sequence <SEQ ID 92>:

```
25

1 MWHWDIILIL LAVGSAAGFI AGLFGVGGGT LIVPVVLWVL DLQGLAQHPY
51 AQHLAVGTSF AVMVFTAFSS MLGQHKKQAV DWKTIFAMMP GMIFGVFAGA
101 LSAKYIPAFG LQIFFILFLT AVAFKTLHTG RQTASRPLPG LPGLTAVSTL
151 FGAMSSWVGI GGGSLSVPFL IHCGFPAHKA IGTSSGLAWP IALSGAISYL
201 VNGLNIAGLP EGSLGFLYLP AVAVLSAATI AFAPLGVKTA HKLSSAKLKE
30
251 SFGIMLLIA GKMLYNLL*
```

Further work revealed the complete gonococcal DNA sequence <SEQ ID 93>:

```
1 ATGTGGCATT GGGACATTAT CTTAATCCTG CTTGCcgtag gcAGTGCGGC
                51 AGGTTTTATT GCCGGCCTGT Tcggtgtagg cggcgGTACG CTGATTGTCC
                    CTGTCGTTTT ATGGGTGCTT GATTTGCAGG GTTTGGCACA ACATCCTTAC
                101
               151 GCGCAACACC TCGCCGTCGG CAcaTccttc gcCGTCATGG TCTTCACCGC
35
                    CTTTTCCAGT ATGTTGGGGC AGCACAAAAA ACAGGCGGTC GACTGGAAAA
                201
                    CCATATTTGC GATGATGCCG GGTATGATAT TCGGCGTATT CGCTGGCGCA
                251
               301 CTCTCCGCAA AATATATCCC CGCGTTCGGG CTTCAAATTT TCTTCATCCT
                351 GTTTTTAACC GCCGTCGCAT TCAAAACACT GCATACCGGT CGTCAGACGG
                    CATCCCGCCC GCTGCCCGGG CTGCCCGGAC TGACTGCGGT TTCCACACTG
40
                401
                    TTCGGCGCAA TGTCGAGCTG GGTCGGCATA GGCGGCGGTT CACTTTCCGT
                451
                501 CCCCTTCTTA ATCCACTGCG GCTTCCCCGC CCATAAAGCC ATCGGCACAT
                551 CATCCGGCCT TGCCTGGCCG ATTGCACTCT CCGGCGCAAT ATCGTATCTG
                601 GTCAACGGTC TGAATATTGC AGGATTGCCC GAAGGGTCGC TGGGCTTCCT
                    TTACCTGCCC GCCGTCGCCG TCCTCAGCGC GGCAACCATT GCCTTTGCCC
45
                    CGCTCGGTGT CAAAACCGCC CACAAACTTT CTTCTGCCAA ACTCAAAGAA
                    TCCTTCGGCA TTATGTTGCT TTTGATTGCC GGAAAAATGC TGTACAACCT
                    GCTTTAA
```

This corresponds to the amino acid sequence <SEQ ID 94; ORF17ng-1>:

```
50 1 MWHWDIILIL LAVGSAAGFI AGLFGVGGGT LIVPVVLWVL DLQGLAQHPY
51 AQHLAVGTSF AVMVFTAFSS MLGQHKKQAV DWKTIFAMMP GMIFGVFAGA
101 LSAKYIPAFG LQIFFILFLT AVAFKTLHTG RQTASRPLPG LPGLTAVSTL
151 FGAMSSWVGI GGGSLSVPFL IHCGFPAHKA IGTSSGLAWP IALSGAISYL
201 VNGLNIAGLP EGSLGFLYLP AVAVLSAATI AFAPLGVKTA HKLSSAKLKE
55 251 SFGIMLLLIA GKMLYNLL*
```

ORF17ng-1 and ORF17-1 show 96.6% identity in 268 aa overlap:

10 20 30 40 50 60 orf17-1.pep MWHWDIILILLAVGSAAGFIAGLFGVGGGTLIVPVVLWVLDLQGLAQHPYAQHLAVGTSF

	orf17ng-1				_		
		10	20	30	40	.50	60
5		70	80	90	100	110	120
	orf17-1.pep	AVMVFTAFSSMLGQ	· -				FILFLT
	orf17ng-1	AVMVFTAFSSMLGQ				· · · · · · · · · · · ·	FILFLT
	•	70	80	90	100	110	120
10				:		4	
		130	140	150	160	170	180
	orf17-1.pep	AVAFKTLHTDPQTA				LSVPFLIHCG	FPAHKA
				111111111:			111111
	orf17ng-1	AVAFKTLHTGRQTA					
15		. 130	140	150	160	170	180
	·	190	200	210	. 220	230	240
	orf17-1.pep	IGTSSGLAWPIALS					
•	orri,-i.beb	IIIIIIIIIIIIIII	IIIIIIIII	1111111111	1111111111	HILLITER	LILLI
20	orf17ng-1	IGTSSGLAWPIALS	GATSYLVNG	INTAGLPEGS	I.GFI.YI.PAVAY	/T.Saattafaf	T.GVKTA
20	OIII/IIg I	190	200	210	220	230	240
		250	260	269			•
	orf17-1.pep	HKLSSAKLKKXFGI	MLLLIAGKM	LYNLLX			
25		1111111111: 111	111111111	111111		•	
	orf17ng-1	HKLSSAKLKESFGI	MLLLIAGKM	LYNLLX			•
	٠, -	250	260				
•		•		•			

In addition, ORF17ng-1 shows significant homology with a hypothetical H.influenzae protein:

```
sp|P44070|Y902_HAEIN HYPOTHETICAL PROTEIN HI0902 pir||G64015 hypothetical protein
30
          HI0902 - Haemophilus influenzae (strain Rd KW20) gi|1573922 (U32772) H. influenzae
          predicted coding region HI0902 [Haemophilus influenzae]Length = 264
           Score = 74 (34.9 bits), Expect = 1.6e-23, Sum P(2) = 1.6e-23
           Identities = 15/43 (34%), Positives = 23/43 (53%)
35
                    55 AVGTSFAVMVFTAFSSMLGQHKKQAVDWKTIFAMMPGMIFGVF 97
          Query:
                       A+GTSFA +V T
                                     S
                                           НK
                                               + W+ + + P ++ VF
                    52 ALGTSFATIVITGIGSAQRHHKLGNIVWQAVRILAPVIMLSVF 94
          Sbjct:
           Score = 195 (91.9 bits), Expect = 1.6e-23, Sum P(2) = 1.6e-23
40
           Identities = 44/114 (38%), Positives = 65/114 (57%)
                    150 LFGAMSSWVGIGGGSLSVPFLIHCGFPAHKAIGTSSGLAWPIALSGAISYLVNGLNIAGL 209
          Query:
                       L G SS GIGGG
                                       VPFL
                                              G
                                                    +AIG+S+
                                                                + +SG S++V+G
                    148 LIGMASSAAGIGGGGFIVPFLTARGINIKQAIGSSAFCGMLLGISGMFSFIVSGWGNPLM 207
           Sbjct:
45
                    210 PEGSLGFLYLPAVAVLSAATIAFAPLGVKTAHKLSSAKLKESFGIMLLLIAGKM 263
          Query:
                        PE SLG++YLPAV ++A + + LG
                                                       KL + LK+ F + L+++A M
                    208 PEYSLGYIYLPAVLGITATSFFTSKLGASATAKLPVSTLKKGFALFLIVVAINM 261
           Sbjct:
```

This analysis, including the homology with the hypothetical *H.influenzae* transmembrane protein, suggests that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 12

50

55 The following partial DNA sequence was identified in N. meningitidis <SEQ ID 95>:

		1	GGAAACGGAT	GGCAGGCAGA	CCCCGAACAT	CCGCTGCTCG	GGCTTTTTGC
		51	CGTCAGTAAT	GTATCGATGA	CGCTTGCTTT	TGTCGGAATA	TGTGCGTTGG
		101	TGCATTATTG	CTTTTCGGGA	ACGGTTCAAG	TGTTTGTGTT	TGCGGCACTG
		151	CTCAAACTTT	ATGCGCTGAA	GCCGGTTTAT	TGGTTCGTGT	TGCAGTTTGT
60	*	201	GCTGATGGCG	GTTGCCTATG	TCCACCGCTG	CGGTATAGAC	CGGCAGCCGC
		251	CGTCAACGTT	CGGCGGCTCG	CAGCTGCGAC	TCGGCGGGTT	GACGGCAGCG

301 351	TTGATGCAGG TCTCGGTACT GGTGCTGCTG CTTTCAGAAA TTGGAAGATA
This corresponds	to the amino acid sequence <seq 96;="" id="" orf18="">:</seq>

5 ...GNGWQADPEH PLLGLFAVSN VSMTLAFVGI CALVHYCFSG TVQVFVFAAL
LKLYALKPVY WFVLQFVLMA VAYVHRCGID RQPPSTFGGS QLRLGGLTAA
LMQVSVLVLL LSEIGR\*

# Further work revealed the complete nucleotide sequence <SEQ ID 97>:

	1	ATGATTTTGC	TGCATTTGGA	TTTTTTGTCT	GCCTTACTGT	ATGCGGCGGT
	51	<b>Փ</b>	CTGATATTCC	GCGCAGGAAT	GTTGCAATGG	TTTTGGGCGA
10		CMARGATCCT	CTCCCTCCCC	ATATCGGTTT	TGGGGGCAAA	GCTGATGCCC
10	101	GIATIAIGCI	GIGGCIGGC	CCCCCCCCCCC	TTGTTCATCC	CCCATTTTTA
	151	GGCATATGGG	GAATGACCCG	CGCCGCGCCC		
	201	CCTGACTTTG	GGCAGCATAT	TTTTTTTCAT	CGGGCATTGG	AACCGGAAAA
	251	CACATGGAAA	CGGATGGCAG	GCAGACCCCG	AACATCCGCT	GCTCGGGCTT
		CUGUIGOUA	CHA AUCUAUC	これがこれではでかか	GCTTTTGTCG	CAATATGTGC
	301	TTTGCCGTCA	GTAATGTATC	GAIGACGCII		omommmo.ccc
15	351	GTTGGTGCAT	TATTGCTTTT	CGGGAACGGT	TCAAGTGTTT	GTGTTTGCGG
13	401	CACTGCTCAA	ACTTTATGCG	CTGAAGCCGG	TTTATTGGTT	CGTGTTGCAG
•	451	TTTGTGCTGA	TGGCGGTTGC	CTATGTCCAC	CGCTGCGGTA	TAGACCGGCA
		1110100101	N CCDDCCCCC	CCTCCCACCT	GCGACTCGGC	GGGTTGACGG
	501	GCCGCCGTCA	ACG11CGGCG	GCICGCAGCI	0000000000	B C B B B B B B C C B
	551	CAGCGTTGAT	GCAGGTCTCG	GTACTGGTGC	TGCTGCTTTC	AGAAATTGGA
20	601	AGATAA				
20	OOT	NONTAN				

This corresponds to the amino acid sequence <SEQ ID 98; ORF18-1>:

```
1 MILLHLDFLS ALLYAAVFLF LIFRAGMLQW FWASIMLWLG ISVLGAKLMP
51 GIWGMTRAAP LFIPHFYLTL GSIFFFIGHW NRKTDGNGWQ ADPEHPLLGL
101 FAVSNVSMTL AFVGICALVH YCFSGTVQVF VFAALLKLYA LKPVYWFVLQ
151 FVLMAVAYVH RCGIDRQPPS TFGGSQLRLG GLTAALMQVS VLVLLLSEIG
201 R*
```

Computer analysis of this amino acid sequence gave the following results:

## Homology with a predicted ORF from N. meningitidis (strain A)

ORF18 shows 98.3% identity over a 116aa overlap with an ORF (ORF18a) from strain A of N.

### 30 meningitidis:

35	orf18.pep	TRAAPLFIPHI 60	FYLTLGSIF 70	FFIGHWNR 80	111111	1111111111	20 AVSNVSMTLA            AVSNVSMTLA 110	1111
		40	5	0	60	70	80	90
	orf18.pep	CALVHYCFSG	rvovrvraa	LLKLYALF	(PVYWFVLQF	<u>VLMAVAYV</u> HR	CGIDROPPST	IIII
40	orf18a	CALVHYCFSX	TVQVFVFAA 130	LLKLYALI 140	XPVYWFVLQF 150	VLMAVAYVHR 160	CGIDROPPST 170	'FGGS
		. 100	11					
46	orf18.pep	OLRIGGITAA	LMQVSVLVL	LLSEIGR	K I			
45	orf18a	QLRLGGLTAA 180	LMOXSVLVI 190	LLSEIGR 200	X			

## The complete length ORF18a nucleotide sequence <SEQ ID 99> is:

	51	TTTTCTGTTT GTATTATGCT GGCATATGGG CCTGACTTTG	CTGATATTCC GTGGCTGGGC GAATGACCCG GGCAGCATAT	GCGCAGGAAT ATATCGGTTT CGCCGCGCCC TTTTTTTCAT	GCCTTACTGT GTTGCAATGG TGGGGGCAAA TTGTTCATCC CGGGCATTGG	TTTTGGGCGA GCTGATGCCC CCCATTTTA AACCGGAAAA
55	251 301 351 401	TTTGCCGTCA	GTAATGTATC TATTGCTTTT	GATGACGCTT CGNGAACGGT	AACATCCTCT GCTTTTGTCG TCAAGTGTTT TTTATTGGTT	GAATATGTGC GTGTTTGCGG

10

451	TTTGTGCTGA	TGGCGGTTGC	CTATGTCCAC	CGCTGCGGTA	TAGACCGGCA
501	GCCGCCGTCA	ACGTTCGGCG	GNTCGCAGCT	GCGACTCGGC	GGGTTGACGG
551	CAGCGTTGAT	GCAGNTCTCG	GTACTGGTGC	TGCTGCTTTC	AGAAATTGGA
601	3C3M33				**

601 AGATA

5 This encodes a protein having amino acid sequence <SEQ ID 100>:

```
1 MILLHLDFLS ALLYAAVFLF LIFRAGMLQW FWASIMLWLG ISVLGAKLMP
51 GIWGMTRAAP LFIPHFYLTL GSIFFFIGHW NRKTDGNGWQ ADPEHPLLGL
101 FAVSNVSMTL AFVGICALVH YCFSXTVQVF VFAALLKLYA LKPYWFVLQ
151 FVLMAVAYVH RCGIDRQPPS TFGGSQLRLG GLTAALMQXS VLVLLLSEIG
201 R*
```

ORF18a and ORF18-1 show 99.0% identity in 201 aa overlap:

	•	10	20	30	40	50	60
	orf18a.pep	MILLHLDFLS	ALLYAAVFLF	LIFRAGMLQWI	FWASIMLWLG:	ISVLGAKLMPG	IWGMTRAAP
_ •		11111111111	1111111111	1111111111		11111111111	
15	orf18-1	MILLHLDFLS	BALLYAAVFLF	LIFRAGMLQWI	FWASIMLWLG:	ISVLGAKLMPG	IWGMTRAAP
		10	20	30	40	50	60
		70			100		120
	orf18a.pep	LFIPHFYLTI	<i>G</i> SIFFFIGHW	NRKTDGNGWQ1	ADPEHPLLGL:	FAVSNVSMTLA	FVGICALVH
20	•			1111111111	* * * * * * * * * * *		
	orf18-1				ADPEHPLLGL:	Favsnvsm <b>tl</b> a	FVGICALVH
•		70	80	90	100	110	120
		130					180
25	orf18a.pep	YCFSXTVQVI	VFAALLKLYA	LKPVYWFVLQ!	FVLMAVAYVH	RCGIDRQPPST	FGGSQLRLG
				<b>{                                    </b>	1111111111	111111111111	11111111
	orf18-1					RCGIDRQPPST	FGGSQLRLG
		130	140	150	160	170	180
30		190	200				
	orf18a.pep	GLTAALMQXS	SVLVLLLSEIG	RX			
		111111111	11111111111	11			
	orf18-1	GLTAALMQV	VLVLLLSEIG	RX			
		190	200				
35	• ,						

Homology with a predicted ORF from N. gonorrhoeae

ORF18 shows 93.1% identity over a 116aa overlap with a predicted ORF (ORF18.ng) from N. gonorrhoeae:

40	orf18.pep	GNGWQADPEHPLLGLFAVSNVSMTLAFVGI	30
	orf18ng		115
	orf18.pep	CALVHYCFSGTVQVFVFAALLKLYALKPVYWFVLQFVLMAVAYVHRCGIDRQPPSTFGGS	90
45	orf18ng	CALVHYCFSGTVQVFVFAALLKLYALKPVYWFVLQFVLMAVAYVHRCGIDRQPPSTFGGS	175
	orf18.pep	QLRLGGLTAALMQVSVLVLLLSEIGR 116	
	orf18ng	QLRLGVLAAMLMQVAVTAMLLAEIGR 201	

50 The complete length ORF18ng nucleotide sequence is <SEQ ID 101>:

	1				GCCTTACTGt	
	. 51	tttTctgTTT	CTGATATTCC	GCGCAGGAAT	GTTGCAATGG	TTTTGGGCGA
	101	GTATTGCGTT'	GTGGCTCGGC	ATCTCGGTTT	TAGGGGTAAA	GCTGATGCCG
_	151				TTGTTCATCC	
55	201	CCTGACTTTG	GGCAGCATAT	TTTTTTTCAT	CGGGTATTGG	AACCGGAAAA
	251	CAGATGGAAA	CGGATGGCAG	GCAGACCCCG	AACATCCGCT	GCTCGGGCTT
	301	TTTGCCGTCA	GTAATGTATC	GATGACGCTT	GCTTTTGTCG	GAATATGTGC
,	351	GTTGGTGCAT	TATTGCTTTT	CGGGAACGGT	TCAAGTGTTT	GTGTTTGCGG
	401				TTTATTGGTT	
60	451	TTTGTATTGA	TGGCGGttgC	CTATGTCCAC	CGCTGCGGTA	TAGACCGGCA
	501	GCCGCCGTCA	ACGTTCGGCG	GTTCGCAGCT	GCGACTCGGC	GTGTTGGCGG

551 CGATGTTGAT GCAGGTTGCG GTAACGGCGA TGCTGCTTGC CGAAATCGGC

601 AGATGA

This encodes a protein having amino acid sequence <SEQ ID 102>:

```
5 MILLHLDFLS ALLYAAVFLF LIFRAGMLQW FWASIALWLG ISVLGVKLMP
51 GMWGMTRAAP LFIPHFYLTL GSIFFFIGYW NRKTDGNGWQ ADPEHPLLGL
101 FAVSNVSMTL AFVGICALVH YCFSGTVQVF VFAALLKLYA LKPVYWFVLQ
151 FVLMAVAYVH RCGIDRQPPS TFGGSQLRLG VLAAMLMQVA VTAMLLAEIG
201 R*
```

This ORF18ng protein sequence shows 94.0% identity in 201 aa overlap with ORF18-1:

10	orf18-1.pep	10 MILLHLDFLSALLYA             MILLHLDFLSALLYA 10	1111111111	1111111111	1 11111111	1:1111:11	11111
15	£19-1 non	70 LFIPHFYLTLGSIFF	80 FIGHWNRKI	90 DGNGWOADPE	100 CHPLLGLFAVS	110 NVSMTLAFVG	120 ICALVH
	orf18-1.pep orf18ng		111:1111				ICALVH
20	012109	70	80	90	100	110 170	120 180
	orf18-1.pep	130 YCFSGTVQVFVFAAI	140 LKLYALKP	150 /YWFVLQFVL!	160 MAVAYVHRCGI		
25	orf18ng	YCFSGTVQVFVFAAI	LKLYALKP	VYWFVLQFVLI 150	MAVAYVHRCG 160	DROPPSTFGG 170	SQLRLG 180
		190	200				
30	orf18-1.pep		11:1111				
	orf18ng	VLAAMLMQVAVTAM 190	LLAEIGRX 200				

Based on this analysis, including the presence of several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 13

35

The following partial DNA sequence was identified in N.meningitidis <SEQ ID 103>:

```
40 51 CGCCAGTGTT TTTACCGCCG CCTCCATCGT CTGCCAGCCCA GGCGAACCCA
101 AGCTCGCCAT GCCCTTCGTA CTCGCCGCGCG CCTTGTCGAT
151 TTGGACAACC NCNTGACCGG ACGCTNAAA AACATCATCA CCACCGTCGC
201 CCTGTTCACC CTCTCCTCGC TCACCGCACA AAGCACCCTC GGCACAGGGC
251 TGCCCTTCAT CCTCGCCATG ACCCTGATGA CTT.CG.CTT CACCATTTTA
45 301 GGCGCGGNCG ...
```

This corresponds to the amino acid sequence <SEQ ID 104; ORF19>:

- 1 MKTPLLKPLL ITSLPVFASV FTAASIVWQL GEPKLAMPFV LGIIAGGLVD 51 LDNXXTGRLK NIITTVALFT LSSLTAQSTL GTGLPFILAM TLMTXXFTIL
- 101 GAX...
- 50 Further work revealed the complete nucleotide sequence <SEQ ID 105>:
  - 1 ATGAAAACCC CACTCCTCAA GCCTCTGCTC ATTACCTCGC TTCCCGTTTT 51 CGCCAGTGTT TTTACCGCCG CCTCCATCGT CTGGCAGCTA GGCGAACCCA
  - 101 AGCTCGCCAT GCCCTTCGTA CTCGGCATCA TCGCCGGCGG CCTTGTCGAT
  - 151 TTGGACAACC GCCTGACCGG ACGGCTGAAA AACATCATCA CCACCGTCGC

```
CCTGTTCACC CTCTCCTCGC TCACGGCACA AAGCACCCTC GGCACAGGGC
                201
                     TGCCCTTCAT CCTCGCCATG ACCCTGATGA CCTTCGGCTT CACCATTTTA
                301
                     GGCGCGGTCG GGCTCAAATA CCGCACCTTC GCCTTCGGTG CACTCGCCGT
                     CGCCACCTAC ACCACACTTA CCTACACCCC CGAAACCTAC TGGCTGACCA
 5
                     ACCCCTTCAT GATTTTATGC GGCACCGTAC TGTACAGCAC CGCCATCCTC
                401
                     CTGTTCCAAA TCGTCCTGCC CCACCGCCCC GTCCAAGAAA GCGTCGCCAA
                451
                     CGCCTACGAC GCACTCGGCG GCTACCTCGA AGCCAAAGCC GACTTCTTCG
                501
                     ACCCCGATGA GGCAGCCTGG ATAGGCAACC GCCACATCGA CCTCGCCATG
                551
                     AGCAACACCG GCGTCATCAC CGCCTTCAAC CAATGCCGTT CCGCCCTGTT
                601
                     TTACCGCCTT CGCGGCAAAC ACCGCCACCC GCGCACCGCC AAAATGCTGC
10
                651
                     GTTACTACTT TGCCGCCCAA GACATACACG AACGCATCAG CTCCGCCCAC
                701
                     GTCGATTATC AGGAAATGTC CGAAAAATTC AAAAACACCG ACATCATCTT
                751
                801
                     CCGCATCCAC CGCCTGCTCG AAATGCAGGG ACAAGCCTGC CGCAACACCG
                     CCCAAGCCCT GCGCGCAAGC AAAGACTACG TTTACAGCAA ACGCCTCGGC
                851
15
                901
                     CGCGCCATCG AAGGCTGCCG CCAATCGCTG CGCCTCCTTT CAGACAGCAA
                     CGACAGTCCC GACATCCGCC ACCTGCGCCG CCTTCTCGAC AACCTCGGCA
                951
               1001
                     GCGTCGACCA GCAGTTCCGC CAACTCCAGC ACAACGGCCT GCAGGCAGAA
                     AACGACCGCA TGGGCGACAC CCGCATCGCC GCCCTCGAAA CCAGCAGCCT
               1051
                     CAAAAACACC TGGCAGGCAA TCCGTCCGCA GCTAAACCTC GAATCAGGCG
               1101
20
                     TATTCCGCCA TGCCGTCCGC CTGTCCCTCG TCGTTGCCGC CGCCTGCACC
               1151
                     ATCGTCGAAG CCCTCAACCT CAACCTCGGC TACTGGATAC TACTGACCGC
               1201
                     CCTTTTCGTC TGCCAACCCA ACTACACCGC CACCAAAAGC CGCGTCCGCC
               1251
                     AGCGCATCGC CGGCACCGTA CTCGGCGTAA TCGTCGGCTC GCTCGTCCCC TACTTCACCC CGTCTGTCGA AACCAAACTC TGGATTGTCA TCGCCAGTAC
               1301
               1351
25
                     CACCCTCTTT TTCATGACCC GCACCTACAA ATACAGTTTC TCCACCTTCT
               1401
                     TCATTACCAT TCAAGCCCTG ACCAGCCTCT CCCTCGCAGG TTTGGACGTA
               1451
                     TACGCCGCCA TGCCCGTACG CATCATCGAC ACCATTATCG GCGCATCCCT
               1501
                     TGCCTGGGCG GCAGTCAGCT ACCTGTGGCC AGACTGGAAA TACCTCACGC
               1551
                     TCGAACGCAC CGCCGCCCTT GCCGTATGCA GCAACGGTGC CTATCTCGAA
               1601
30
               1651
                     AAAATCACCG AACGCCTCAA AAGCGGCGAA ACCGGCGACG ACGTCGAATA
                     CCGCGCCACC CGCCGCCGC CCCACGAACA CACCGCCGCC CTCAGCAGCA
               1701
                     CCCTTTCCGA CATGAGCAGC GAACCCGCAA AATTCGCCGA CAGCCTGCAA
               1751
                     CCCGGCTTTA CCCTGCTCAA AACCGGCTAC GCCCTGACCG GCTACATCTC
               1801
               1851
                     CGCCCTCGGC GCATACCGCA GCGAAATGCA CGAAGAATGC AGCCCCGACT
35
                     TTACCGCACA GTTCCACCTC GCCGCCGAAC ACACCGCCCA CATCTTCCAA
               1901
                     CACCTGCCCG AAACCGAACC CGACGACTTT CAGACAGCAC TGGATACACT
               1951
                     GCGCGGCGAA CTCGACACCC TCCGCACCCA CAGCAGCGGA ACACAAAGCC
                     ACATCCTCCT CCAACAGCTC CAACTCATCG CCCGACAGCT CGAACCCTAC
               2051
                     TACCGCGCCT ACCGCCAAAT TCCGCACAGG CAGCCCCAAA ATGCAGCCTG
               2101
40
               2151
```

This corresponds to the amino acid sequence <SEQ ID 106; ORF19-1>:

```
MKTPLLKPLL ITSLPVFASV FTAASIVWQL GEPKLAMPFV LGIIAGGLVD
                      LDNRLTGRLK NIITTVALFT LSSLTAQSTL GTGLPFILAM TLMTFGFTIL
                      GAVGLKYRTF AFGALAVATY TTLTYTPETY WLTNPFMILC GTVLYSTAIL
45
                      LFQIVLPHRP VQESVANAYD ALGGYLEAKA DFFDPDEAAW IGNRHIDLAM
                 151
                      SNTGVITAFN QCRSALFYRL RGKHRHPRTA KMLRYYFAAQ DIHERISSAH
VDYQEMSEKF KNTDIIFRIH RLLEMQGQAC RNTAQALRAS KDYVYSKRLG
                 201
                 251
                 301
                      RAIEGCROSL RLLSDSNDSP DIRHLRRLLD NLGSVDQQFR QLQHNGLQAE
                      NDRMGDTRIA ALETSSLKNT WQAIRPQLNL ESGVFRHAVR LSLVVAAACT
                 351
50
                 401
                      IVEALNLNLG YWILLTALFV CQPNYTATKS RVRQRIAGTV LGVIVGSLVP
                      YFTPSVETKL WIVIASTTLF FMTRTYKYSF STFFITIQAL TSLSLAGLDV
                 451
                      YAAMPVRIID TIIGASLAWA AVSYLWPDWK YLTLERTAAL AVCSNGAYLE
                 501
                      KITERLKSGE TGDDVEYRAT RRRAHEHTAA LSSTLSDMSS EPAKFADSLQ
                      PGFTLLKTGY ALTGYISALG AYRSEMHEEC SPDFTAQFHL AAEHTAHIFQ
                 601
55
                      HLPETEPDDF QTALDTLRGE LDTLRTHSSG TQSHILLQQL QLIARQLEPY YRAYRQIPHR QPQNAA*
                 651
```

Computer analysis of this amino acid sequence gave the following results:

Homology with predicted transmenbrane protein YHFK of *H. influenzae* (accession number P44289) ORF19 and YHFK proteins show 45% as identity in 97 as overlap:

```
60 orf19 6 LKPLLITSLPVFASVFTAASIVWQLGEPKLAMPFVLGIIAGGLVDLDNXXTGRLKNIITT 65
L +I+++PVF +V AA +W +MP +LGIIAGGLVDLDN TGRLKN+ T
YHFK 5 LNAKVISTIPVFIAVNIAAVGIWFFDISSQSMPLILGIIAGGLVDLDNRLTGRLKNVFFT 64
```

orf19 66 VALFTLSSLTAQSTLGTGLPFILAMTLMTXXFTILGA 102 + F++SS Q +G + +I+ MT++T FT++GA YHFK 65 LIAFSISSFIVQLHIGKPIQYIVLMTVLTFIFTMIGA 101

5 Homology with a predicted ORF from N.meningitidis (strain A)

ORF19 shows 92.2% identity over a 102aa overlap with an ORF (ORF19a) from strain A of N. meningitidis:

		10	20	30	40	50	60
	orf19.pep	MKTPLLKPLLITSL	PVFASVFTAA	SIVWQLGEPK 	LAMPFVLGII	AGGLVDLDNX	XTGRLK
10	orf19a	MKTPPLKPLLITSL		SIVWOLGEPK	LAMPEVLGII 40	AGGLVDLDNR	LTGRLK 60
		10	20	30	40	20	00
		70	80	90	100		
15	orf19.pep	NIITTVALFTLSSL	TAQSTLGTGL	PFILAMTLMI	XXFTILGAX		
	orf19a	NIIATVALFTLSSL	VAOSTLGTGL	PFILAMTLMI	FGFTIMGAV	<b>SLKYRTFAFGA</b>	LAVATY
	OLLIDA	70	80	90	100	110	120
20	orf19a	TTLTYTPETYWLTN	PFMILCGTVI	YSTAILLFQ	[LPHRPVQE]	NVANAYEALGS	YLEAKA
20	ULLIJU	130	140	150	160	170	180

The complete length ORF19a nucleotide sequence <SEQ ID 107> is:

		_				
	1	ATGAAAACCC	CACCCCTCAA	GCCTCTGCTC	ATTACCTCGC	TTCCCGTTTT
	51	CCCCACTCTC	TTTACCGCCG	CCTCCATCGT	CTGGCAGCTG	GGCGAACCCA
25	101	A COTTO COO TO	GCCCTTCGTA	CTCGGCATCA	TCGCTGGCGG	CCTGGTCGAT
25	151	TTCCACAACC	GCCTGACCGG	ACGGCTGAAA	AACATCATCG	CCACCGTCGC
	201	CCTGTTCACC	CTCTCCTCAC	TTGTCGCGCA	AAGCACCCTC	GGCACAGGTT
	251	mCCCNTTCNT	CCTCGCCATG	ACCCTGATGA	CTTTCGGCTT	TACCATCATG
	301	CCCCCCCCCCCC	CCCTGAAATA	CCGCACCTTC	GCCTTCGGCG	CACTCGCCGT
20	351	CCCCACCTAC	ACCACACTTA	CCTACACCCC	CGAAACCTAC	TGGCTGACCA
30	401	እ CCCCጥጥጥ እጥ	CATTCTCTCC	GGAACCGTAC	TGTACAGCAC	CGCCATCATC
	451	CMCMMCCNNN	TCATCCTGCC	CCACCGCCCC	GTTCAAGAAA	ACGTCGCCAA
	501	CCCCTACCAA	GCACTCGGCA	GCTACCTCGA	AGCCAAAGCC	GACTTTTTCG
		A TOCOCO A COA	ACCCGAATGG	ATAGGCAACC	GCCACATCGA	CCTCGCCATG
25	551 601	ACCAACACCC	CCCTCATCAC	CGCCTTCAAC	CAATGCCGTT	CCGCCCTGTT
35	651	ጥም አ ር ር ር ር ር ር ጥጥ	CCCCCCAAAC	ACCGCCACCC	GCGCACCGCC	AAAATGCTGC
	701	ርር መስር መስር ማጥ	CCCCCCCAA	GACATACACG	AACGCATCAG	CTCCGCCCAC
	751	<b>で中ででみで甲みでで</b>	AAGAGATGTC	CGAAAAATTC	AAAAACACCG	ACATCATCTT
	801	CCCCATCCAC	CGCCTGCTCG	AAATGCAGGG	ACAAGCCTGC	CGCAACACCG
40	851	CCCNACCCCT	CCCCCCAAGC	AAAGACTACG	TTTACAGCAA	ACGCCTCGGC
40	901	CCCCCCATCG	AAGGCTGCCG	CCAATCGCTG	CGCCTCCTTT	CAGACAGCAA
	951	CCACAATCCC	GACATCCGCC	ACCTGCGCCG	CCTTCTCGAC	AACCTCGGCA
	1001	CCCTCCACCA	CCAGTTCCGC	CAACTCCAGC	ACAACGGCCT	GCAGGCAGAA
		A D C C D C C C C D	TEGECENCAC	CCGCATCGCC	GCCCTCGAAA	CCGGCAGCCT
4.5	1051 1101	CANANACACC	TCCCAGCAA	TCCGTCCGCA	GCTAAACCTC	GAATCAGGCG
45		TATATACACC	TECCETCESC	CTGTCCCTTG	TCGTTGCCGC	CGCCTGCACC
	1151 1201	AMCCACCAVC	CCCTCAACCT	CAACCTCGGC	TACTGGATAC	TACTGACCGC
		CCTTTTCCTC	TGCCAACCCA	ACTACACCGC	CACCAAAAGC	CGCGTCCGCC
	1251 1301	ACCCCATCCC	CGCCACCGTA	CTCGGCGTAA	TCGTCGGCTC	GCTCGTCCCC
<b>~</b> ^	1351	MGCGCATCGC	CCTCCGTCGA	AACCAAACTC	TGGATCGTCA	TCGCCAGTAC
50		CACCCCCCCCC	TTTCATGACCC	GCACCTACAA	ATACAGCTTC	TCGACATTTT
	1401	TCAUCCICIII	TCAAGCCCTG	ACCAGCCTCT	CCCTCGCAGG	GTTGGACGTA
	1451 1501	TONICACCAI	TGCCCGTACG	CATCATCGAC	ACCATTATCG	GCGCATCCCT
		TACGCCGCCA	CCACTCACCT	ACCTGTGGCC	AGACTGGAAA	TACCTCACGC
	1551	TGCC1GGGCG	CCCCCCCCCTT	GCCGTATGCA	GCAACGGCGC	CTATCTCGAA
55	1601	T CGAACGCAC	AACGCCTCAA	AAGCGGCGAA	ACCGGCGACG	ACGTCGAATA
	1651	AAAATCACCC	CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	CCCACGAACA	CACCGCCGCC	CTCAGCAGCA
	1701	CCGCGCCACC	CATCACCACC	CAACCCGCAA	AATTCGCCGA	CAGCCTGCAA
	1751		CCCCCCCCCA	AACCGGCTAC	GCCCTGACCC	GCTACATCTC
	1801		CCCIGCION	GCGAAATGC	CGAAGAATGO	AGCCCCGACT
60	1851	mma cccca ca	CULTACOGO.	CCCCCCGAA	ACACCGCCC	CATCTTCCAA
	1901	TTACCGCACA	A GIICCHCCIC	CCACCACTT	CAGACAGCAC	TGGATACACT
	1951	CACCTGCCCC	AMMCCGMAC(	TORCOROTT	CAGCAGCGG	ACACAAAGCC
	2001		T CLOGHONCO	CAACTCATC	CCCGGCAGC'	CGAACCCTAC
	2051		T ACCCACAAA'	T TOTOLOGICATOR	CAGCCCCAA	A ACGCAGCCTG
65	2101		ACCOMUMAN.	LOCUMUNU		
	2151	. A				

# This encodes a protein having amino acid sequence <SEQ ID 108>:

	1	MKTPPLKPLL	ITSLPVFASV	FTAASIVWQL	GEPKLAMPFV	LGIIAGGLVD
	51	LDNRLTGRLK	NIIATVALFT	LSSLVAQSTL	GTGLPFILAM	TLMTFGFTIM
	101	GAVGLKYRTF	AFGALAVATY	TTLTYTPETY	WLTNPFMILC.	GTVLYSTAII
5	151	LFQIILPHRP	VQENVANAYE	ALGSYLEAKA	DFFDPDEAEW	IGNRHIDLAM
•	201				KMLRYYFAAQ	
	251	VDYQEMSEKF	KNTDIIFRIH	RLLEMQGQAC	RNTAQALRAS	KDYVYSKRLG
	301	RAIEGCRQSL	RLLSDSNDNP	DIRHLRRLLD	NLGSVDQQFR	QLQHNGLQAE
. •	351	NDRMGDTRIA	ALETGSLKNT	WQAIRPQLNL	ESGVFRHAVR	LSLVVAAACT
10	401	IVEALNLNLG	YWILLTALFV	CQPNYTATKS	RVRQRIAGTV	LGVIVGSLVP
	451	YFTPSVETKL	WIVIASTTLE	FMTRTYKYSF	STFFITIQAL	TSLSLAGLDV
	501	YAAMPVRIID	TIIGASLAWA	AVSYLWPDWK	YLTLERTAAL	AVCSNGAYLE
	551	KITERLKSGE	TGDDVEYRAT	RRRAHEHTAA	LSSTLSDMSS	EPAKFADSLQ
	601	PGFTLLKTGY	ALTGYISALG	AYRSEMHEEC	SPDFTAQFHL	AAEHTAHIFQ
15	651	HLPETEPDDF	QTALDTLRGE	LDTLRTHSSG	TQSHILLQQL	QLIARQLEPY
••	701	YRAYRQIPHR	QPQNAA*	-		•

ORF19a and ORF19-1 show 98.3% identity in 716 aa overlap:									
	orf19a.pep	10 MKTPPLKPLLITSLP	20 VFASVETA	30 ASTVWOLGEPKI	40	50	60 TGRLK		
20	OIII3a.pep		11111111	ШНЁНШ	111111111	E			
	orf19-1	MKTPLLKPLLITSLP		ASIVWQLGEPKL 30			LTGRLK 60		
		10	20	30	40	50	ėυ		
	•	70	- 80	90	100	110	120		
25	orf19a.pep	NIIATVALFTLSSLV	_						
	orf19-1	NIITTVALFTLSSLT							
		70	. 80	90	100	110	120		
30		130	140	150	160	170	180		
	orf19a.pep	TTLTYTPETYWLTNP	FMILCGT	/LYSTAIILFQII	LPHRPVQE	NVANAYEALGS:			
	<b>510</b> 1								
	orf19-1	130	140	150	160	170	180		
35									
	orf19a.pep	190 DFFDPDEAEWIGNRH	200 מאמוחדי	210 PGVTTA FNOCRSI	220	230 нвиретакмте	240 YYFAAO		
	OIII3a.pep								
40	orf19-1	DFFDPDEAAWIGNRH							
40	•	190	200	210	220	230	240		
	•	250	260	270	280	290	300		
	orf19a.pep	DIHERISSAHVDYQE							
45	orf19-1	DIHERISSAHVDYQE							
		250	260	270	280	290	300		
		310	320	330	340	350	360		
	orf19a.pep	RAIEGCRQSLRLLSD		-					
50		PATECOROSIDATO							
	orf19-1	RAIEGCROSLRLLSD 310	320 · 320	330	иродеково 340	HNGLQAENDRM 350	360		
	•								
55	orf19a.pep	370 ALETGSLKNTWOAIF	380	390 CVFRHAVR1.S1.VV	400 ממזיים ממני	410	420 1.TA1.EV		
33	Offisa.pep								
	orf19-1	ALETSSLKNTWQAIF							
	· ·	370	380	390	400	410	420		
60		430	440	450	460	470	480		
	orf19a.pep	CQPNYTATKSRVRQF							
	orf19-1	COPNYTATKSRVRQ							
		430	440	450	460	470	480		
65		490	500	510	520	530	540		
	orf19a.pep	STFFITIQALTSLS	LAGLDVYA	AMPVRIIDTIIG	ASLAWAAVS	YLWPDWKYLTL	ERTAAL		
		ППППППППППППППППППППППППППППППППППППППП	шш	111111111111	1111111111	111111111111111111111111111111111111111	111111		

PCT/IB98/01665 WO 99/24578

4	4	^
- 1	1	2-

orf19-1	STFFITIQALTSL	SLAGLDVYAAN	PVRIIDTII	SASLAWAAVSY	LWPDWKYLT!	LERTAAL
	490	500	510	520	530	540
	550	560	570	580	590	600
orf19a.pep	AVCSNGAYLEKIT	ERLKSGETGDI	OVEYRATRRR <i>I</i>	AHEHTAALSST	'LSDMSSEPAI	KFADSLQ
orf19-1	AVCSNGAYLEKIT	ERLKSGETGDI	OVEYRATRRRA	AHEHTAALSST	LSDMSSEPA	KFADSLQ
	550	560	570	580	590	600
	610	620	630	640	650	660
orf19a.pep	PGFTLLKTGYALT	GYISALGAYRS	SEMHEECSPDI	FTAQFHLAAEH	ITAHI FQHLPI	ETEPDDF
	11111111111111	11111111111		11111111111	нийи	1111111
orf19-1	PGFTLLKTGYALT	GYISALGAYRS	SEMHEECSPDI	TAOFHLAAEH	ITAHI FOHLP	ETEPDDF
	610	620	630	640	650	660
					,	
	670	680	690	700	710	
orf19a.pep	QTALDTLRGELDT	LRTHSSGTQSI	HILLQQLQLI	arqlepyyray	(RQIPHRQPQ)	XAAN
				[		1111
orf19-1	QTALDTLRGELDT	LRTHSSGTQSI	HILLQQLQLIA	ARQLEPYYRAY	RQIPHRQPQ	XAAN
	670	680	690	700	710	
	orf19a.pep orf19-1 orf19a.pep orf19-1 orf19a.pep	490  550  orf19a.pep AVCSNGAYLEKIT                orf19-1 AVCSNGAYLEKIT 550  610  orf19a.pep PGFTLLKTGYALT                orf19-1 PGFTLLKTGYALT 610  670  orf19a.pep QTALDTLRGELDT               orf19-1 QTALDTLRGELDT	490 500  550 560  orf19a.pep AVCSNGAYLEKITERLKSGETGDE	490 500 510  550 560 570  orf19a.pep AVCSNGAYLEKITERLKSGETGDDVEYRATRRA	490 500 510 520  550 560 570 580  orf19a.pep AVCSNGAYLEKITERLKSGETGDDVEYRATRRAHEHTAALSST	490 500 510 520 530  550 560 570 580 590  orf19a.pep AVCSNGAYLEKITERLKSGETGDDVEYRATRRAHEHTAALSSTLSDMSSEPA

# Homology with a predicted ORF from N.gonorrhoeae

ORF19 shows 95.1% identity over a 102aa overlap with a predicted ORF (ORF19.ng) from N. gonorrhoeae:

```
25
        orf19.pep
                 MKTPLLKPLLITSLPVFASVFTAASIVWQLGEPKLAMPFVLGIIAGGLVDLDNXXTGRLK
                                                                60
                 MKTPLLKPLLITSLPVFASVFTAASIVWQLGEPKLAMPFVLGIIAGGLVDLDNRLTGRLK
        orf19ng
                 NIITTVALFTLSSLTAQSTLGTGLPFILAMTLMTXXFTILGAX
                                                               103
        orf19.pep
30
                 orf19ng
                 NIIATVALFTLSSLTAQSTLGTGLPFILAMTLMTFGFTILGAVGLKYRTFAFGALAVATY
```

An ORF19ng nucleotide sequence <SEQ ID 109> is predicted to encode a protein having amino acid sequence <SEQ ID 110>:

```
1 MKTPLLKPLL ITSLPVFASV FTAASIVWQL GEPKLAMPFV LGIIAGGLVD
35
                        LDNRLTGRLK NIIATVALFT LSSLTAQSTL GTGLPFILAM TLMTFGFTIL
                        GAVGLKYRTF AFGALAVATY TTLTYTPETY WLTNPFMILC GTVLYSTAII
                  101
                       LFQIILPHRP VQESVANAY<u>E ALGGYLEAKA DFFDP</u>DEAAW IGNRHIDLAM SNTGVITAFN QCRSALFYRL RGKHRHPRTA KMLRYYFAAQ DIHERISSAH
                  151
                  201
                  251
                        VDYQEMSEKF KNTDIIFRIR RLLEMQGQAC RNTAQAIRSG KDYVYSKRLG
40
                        RAIEGCROSL RLLSDGNDSP DIRHLSRLLD NLGSVDQQFR QLRHSDSPAE
                  301
                  351
                        NDRMGDTRIA ALETGSFKNT
```

#### Further work revealed the complete nucleotide sequence <SEQ ID 111>:

	1	ATGAAAACCC	CACTCCTCAA	GCCTCTGCTC	ATTACCTCGC	TTCCCGTTTT
	51	CGCCAGTGTC	TTTACCGCCG	CCTCCATCGT	CTGGCAGCTA	GGCGAACCCA
45	101	AGCTCGCCAT	GCCCTTCGTA	CTCGGCATCA	TCGCCGGCGG	CCTGGTCGAT
	151	TTGGACAACC	GCCTGACCGG	ACGGCTGAAA	AACATCATCG	CCACCGTCGC
	201	CCTGTTTACC	CTCTCCTCGC	TCACGGCGCA	AAGCACCCTC	GGCACAGGGC
	251	TGCCCTTCAT	CCTCGCCATG	ACCCTGATGA	CCTTCGGCTT	TACCATTTTA
_	301	GGCGCGGTCG	GGCTGAAATA	CCGCACCTTC	GCCTTCGGCG	CACTCGCCGT
50	351	CGCCACCTAC	ACCACGCTTA	CCTACACCCC	CGAAACCTAC	TGGCTGACCA
	401	ACCCCTTCAT	GATTTTATGC	GGCACCGTAC	TGTACAGCAC	CGCCATCATC
	451	CTGTTCCAAA	TCATCCTGCC	CCACCGCCCC	GTCCAAGAAA	GCGTCGCCAA
	501	TGCCTACGAA	GCACTCGGCG	GCTACCTCGA	AGCCAAAGCC	GACTTCTTCG
	551	ACCCCGATGA	GGCAGCCTGG	ATAGGCAACC	GCCACATCGA	CCTCGCCATG
55	601	AGCAACACCG	GCGTCATCAC	CGCCTTCAAC	CAATGCCGTT	CCGCCCTGTT
	651	TTACCGTTTG	CGCGGCAAAC	ACCGCCACCC	GCGCACCGCC	AAAATGCTGC
	701	GCTACTACTT	CGCCGCCCAA	GACATCCACG	AACGCATCAG	CTCCGCCCAC
	751	GTCGACTACC	AAGAGATGTC	CGAAAAATTC	AAAAACACCG	ACATCATCTT
	801		CGCCTGCTCG			
60	851	CCCAAGCCAT	CCGGTCGGGC	AAAGACTAcg	<b>tTTACAGCAA</b>	ACGCCTCGGA
	901	CGCGCCATcg	aaggctgCCG	CCAGTCGCtg	cgcctCCTTt	cagacggcaA
•	951	CGACAGTCCC	GACATCCGCC	ACCTGAGCCG	CCTTCTCGAC	AACCTCGgca

	1051 Aac	Togacca gcagtTCcgc gaccgca tgggcgacaC aaaCAcc tggcaggCAA	CCGCATCGCC	GCCCtcgaaa	ccggcagctT
		CCGCCA TGCCGTCCGC	CTGTCCCTCG	TCGTTGCCGC	CGCCTGCACC
5	1201 ATC	GTCgaag cCCTCAACCT	CAACCTCGGC	TACTGGATAC	TGCTGACCGC
	1251 CCT 1301 AAC	ITTCGTC TGCCAACCCA GCATCGC CGGCACCGTA	ACTACACCGC	CACCAAAAGC	CGCGTGTACC
	1301 AAC	TCACCC CCTCCGTCGA	AACCAAACTC	TCGTCGGCTC	TCCCCCTAC
		CCTGTTC TTCATGACCC			
10	1451 TCA	TCACCAT TCAGGCACTG	ACCAGCCTCT	CCCTCGCAGG	TTTGGACGTA
		GCCGCCA TGCCCGTGCG			
		CTGGGCG GCGGTCAGCT AACGCAC CGCCGCCCTT			
		ATTGCCG AACGCCTCAA			
15		CATCACC CGCCGCCGCG			
		ITTCCGA CATGAGCAGC			
		GGCTTTA CCCTGCTCAA			
	1851 CGC 1901 TTA	CCTCGGC GCATACCGCA CCGCACA GTTCCACCTT	GCGAAATGCA	CGAAGAATGC	AGCCCCGACT
20	1951 CAC	CTGCCCG ACATGGGACC	CGACGACTTT	CAGACGGCAT	TGGATACACT
20	2001 GCG	CGGCGAA CTCGGCACCC	TCCGCACCCG	CAGCAGCGGA	ACACAAAGCC
	2051 ACA	TCCTCCT CCAACAGCTC	CAACTCATCG	CccgGCAACT	CGAACCCTAC
		CGCGCCT ACCGACAAAT	TCCGCACAGG	CAGCCCCAAA	ACGCAGCCTG
*	2151 A				
25	This corresponds to	the amino acid seque	ence <seq i<="" th=""><th>D 112; ORF</th><th>19ng-1&gt;:</th></seq>	D 112; ORF	19ng-1>:
	1 <u>MKT</u>	PLLKPLL ITSLPVFASV	FTAASIVWQL	GEPKLAMPFV	LGIIAGGLVD
		RLTGRLK NIIATVALFT			
	101 <u>GAV</u> 151 LFO	GLKYRTF AFGALAVATY	TTLTYTPETY	WLTNPFMILC	GTVLYSTAII
30	201 SNT	IILPHRP VQESVANAYE GVITAFN QCRSALFYRL	RGKHRHPRTA	KMI.RYYFAAO	DIHEDISSAH
50	251 VDY	QEMSEKF KNTDIIFRIR	RLLEMOGOAC	RNTAOAIRSG	KDYVYSKRI G
	301 RAI	EGCRQSL RLLSDGNDSP	DIRHLSRLLD	NLGSVDQQFR	QLRHSDSPAE
		MGDTRIA ALETGSFKNT			
25	401 <u>IVE</u>	ALNLNLG YWILLTALFV	CQPNYTATKS	RVYQRIAGTV	LGVIVGSLVP
35	451 <u>YF</u> T 501 <u>YA</u> A	PSVETKL WIVIAGTTLF MPVRIID TIIGASLAWA	MIRTIKISE	STEFITIQAL	TSLSLAGLDV
	551 KIA	ERLKTGE TGDDIEYRIT	RRRAHEHTAA	LSSTLSDMSS	EPAKFADSI O
	601 PGF	TLLKTGY ALTGYISALG	AYRSEMHEEC	SPDFTAQFHL	AAEHTAHIFO
4.0	651 HLP	DMGPDDF QTALDŤLRGE	LGTLRTRSSG	TQSHILLQQL	QLIARQLEPY
40	701 YRA	YRQIPHR QPQNAA*			
	ORF19ng-1 and OR	F19-1 show 95.5% i	dentity in 71	б aa overlap:	
	•	.10	20	30	40 50 60
	orf19-1.pep	MKTPLLKPLLITSLPV	FASVFTAASIV	WQLGEPKLAMP	<b>FVLGIIAGGLVDLDNRLTGRLK</b>
45	orf19ng-1	MYMDIIYDIITMOIDU		HOLORDKI MAR	
43	Offishg-1	10	20		FVLGIIAGGLVDLDNRLTGRLK 40 50 60
			-4		- 30
	-:-	70	80		00 110 120
50	orf19-1.pep	NIITTVALFTLSSLTA	QSTLGTGLPFI:	LAMTLMTFGFT:	ILGAVGLKYRTFAFGALAVATY
50	orf19ng-1	וון וון וון וון וון א			
	OIII9IIg-I	70	SO 80		DO 110 120
					110 120
		130			60 170 180
55	orf19-1.pep	TTLTYTPETYWLTNPF	MILCGTVLYST	AILLFQIVLPHI	RPVQESVANAYDALGGYLEAKA
			HILLOCOMIE VON	!   :         :	
	orf19ng-1	130			RPVQESVANAYEALGGYLEAKA 60 170 180
		150	140	150 1	60 170 180
60		190			20 230 240
	orf19-1.pep	DFFDPDEAAWIGNRHI	DLAMSNTGVIT	AFNOCRSALFY	RLRGKHRHPRTAKMLRYYFAAC
			11111111111		1111111111111111111111
	orf19ng-1	DFFDPDEAAWIGNRHI 190			RLRGKHRHPRTAKMLRYYFAAC
65	•	130	200	210 2:	20 230 240
55		250			80 290 300
	orf19-1.pep	DIHERISSAHVDYQEM			ACRNTAQALRASKDYVYSKRLG

WO 99/24578 PCT/IB98/01665

	orf19ng-1	
5	orf19-1.pep	310 320 330 340 350 360 RAIEGCRQSLRLLSDSNDSPDIRHLRRLLDNLGSVDQQFRQLQHNGLQAENDRMGDTRIA
10	orf19ng-1	RAIEGCROSLRLLSDGNDSPDIRHLSRLLDNLGSVDQOFRQLRHSDSPAENDRMGDTRIA 310 320 330 340 350 360
10	orf19-1.pep	370 380 390 400 410 420 ALETSSLKNTWQAIRPQLNLESGVFRHAVRLSLVVAAACTIVEALNLNLGYWILLTALFV
15	orf19ng-1	
	orf19-1.pep	430 440 450 460 470 480 CQPNYTATKSRVRQRIAGTVLGVIVGSLVPYFTPSVETKLWIVIASTTLFFMTRTYKYSF
20	orf19ng-1	
25	orf19-1.pep	490 500 510 520 530 540 STFFITIQALTSLSLAGLDVYAAMPVRIIDTIIGASLAWAAVSYLWPDWKYLTLERTAAL
	orf19ng-1	STFFITIOALTSLSLAGLDVYAAMPVRIIDTIIGASLAWAAVSYLWPDWKYLTLERTAAL 490 500 510 520 530 540
30	orf19-1.pep	550 560 570 580 590 600 AVCSNGAYLEKITERLKSGETGDDVEYRATRRAHEHTAALSSTLSDMSSEPAKFADSLQ
25	01113119	550 560 570 580 590 600
35	orf19-1.pep	610 620 630 640 650 660 PGFTLLKTGYALTGYISALGAYRSEMHEECSPDFTAQFHLAAEHTAHIFQHLPETEPDDF
40	orf19ng-1	PGFTLLKTGYALTGYISALGAYRSEMHEECSPDFTAQFHLAAEHTAHIFQHLPDMGPDDF 610 620 630 640 650 660
.•	orf19-1.pep	670 680 690 700 710  QTALDTLRGELDTLRTHSSGTQSHILLQQLQLIARQLEPYYRAYRQIPHRQPQNAAX
45	orf19ng-1	QTALDTLRGELGTLRTRSSGTQSHILLQQLQLIARQLEPYYRAYRQIPHRQPQNAAX 670 680 690 700 710

In addition, ORF19ng-1 shows significant homology to a hypothetical gonococcal protein previously entered in the databases:

50	(AJ00242) Score =	3) hy 1512	R2_NEIGO HYPOTHETICAL 45.5 KD PROTEIN (ORF2) gnl PID e /pothetical protein [Neisseria gonorrh] Length = 417 2 (705.6 bits), Expect = 5.3e-203, P = 5.3e-203 = 301/326 (92%), Positives = 306/326 (93%)	1154438
	Query:	307	${\tt RQSLRLLSDGNDSPDIRHLSRLLDNLGSVDQQFRQLRHSDSPAENDRMGDTRIAALETGS}$	366
			RQSLRLLSDGNDS DIRHLSRLLDNLGSVDQQFRQLRHSDSPAENDRMGDTRIAALETGS	
55	Sbjct:	1	${\tt RQSLRLLSDGNDSXDIRHLSRLLDNLGSVDQQFRQLRHSDSPAENDRMGDTRIAALETGS}$	60
	Query:	367	FKNTWQAIRPQLNLESCVFRHAVRLSLVVAAACTIVEALNLNLGYWILLTALFVCQPNYT	426
			FKNTWQAIRPQLNLES VFRHAVRLSLVVAAACTIVEALNLNLGYWILLT LFVCQPNYT	
60	Sbjct:	61	${\tt FKNTWQAIRPQLNLESGVFRHAVRLSLVVAAACTIVEALNLNLGYWILLTRLFVCQPNYT}$	120
00	Query:	427	ATKSRVYQRIAGTVLGVIVGSLVPYFTPSVETKLWIVIAGTTLFFMTRTYKYSFSTFFIT	486
	- •		ATKSRVYQRIAGTVLGVIVGSLVPYFTPSVETKLWIVIAGTTLFFMTRTYKYSFSTFFIT	100
	Sbjct:	121	${\tt ATKSRVYQRIAGTVLGVIVGSLVPYFTPSVETKLWIVIAGTTLFFMTRTYKYSFSTFFIT}$	180
65	Query:	487	IQALTSLSLAGLDVYAAMPVRIIDTIIGASLAWAAVSYLWPDWKYLTLERTAALAVCSSG	546
			IQALTSLSLAGLDVYAAMPVRIIDTIIGASLAWAAVSYLWPDWKYLTLERTAALAVCSSG	540
	Sbjct:	181	IQALTSLSLAGLDVYAAMPVRIIDTIIGASLAWAAVSYLWPDWKYLTLERTAALAVCSSG	240

```
Query: 547 TYLQKIAERLKTGETGDDIEYRITRRRAHEHTAALSSTLSDMSSEPAKFADSLQPGFTLL 606
TYLQKIAERLKTGETGDDIEYRITRRRAHEHTAALSSTLSDMSSEPAKFAD+ P
Sbjct: 241 TYLQKIAERLKTGETGDDIEYRITRRRAHEHTAALSSTLSDMSSEPAKFADTCNPALPCS 300

5 Query: 607 KTGYALTGYISALGAYRSEMHEECSP 632
K ALTGYISALG ++ + +P
Sbjct: 301 KPATALTGYISALGHTAAKCTKNAAP 326
```

Based on this analysis, including the presence of several putative transmembrane domains in the gonococcal protein (the first of which is also seen in the meningococcal protein), and on homology with the YHFK protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 14

10

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 113>:

```
15
                     ATGAATATGC TGGGAGCTTT GGCAAAAGTC GGCAGCCTGA CGATGGTGTC
                 51
                     GCGCGTTTTG GGATTTGTGC GCGATACGGT CATTGCGCGG GCATTCGGCG
                     CGGGTATGGC GACGGATGCG TTTTTTGTCG CGTTCAAACT GCCCAACCTG
                101
                     CTTCGCCGCG TGTTTGCGGA GGGGGCGTTT GCCCAAGCGT TTGTGCCGAT
                151
                     TTTGGCGGAA TACAAGGAAA CGCGTTCAAA AGAGGCGG.C GAAGCCTTTA
                201
20
                251
                     TCCGCCATGT GGCGGGGATG CTGTCGTTTG TACTGGTTAT CGTTACCGCG
                     CTGGGCATAC TTGCCGCGCC TTGGGTGATT TATGTTTCCG CACCCGAGTT
                301
                351
                     TTGCCCAAGA TGCCGACAAA TTTCAGCTCT CCATCGATTT GCTGCGGATT
                     ACGTTTCCTT ATATATTATT GATTTCCCTG TCTTCATTTG TCGGCTCGGT
                401
                     ACTCAATTCT TATCATAAGT TCGGCATTCC GGCGTTTACG CCAC.GTTTC
                451
25
                     TGAACGTGTC GTTTATCGTA TTCGCGCTGT TTTTCGTGCC GTATTTCGAT
                501
                551
                     CCGCCCGTTA CCGCGCYGGC GTGGGCGGTC TTTGTCGGCG GCATTTTGCA
                601
                     ACTCGrmTTC CAACTGCCCT GGCTGGCGAA ACTGGGCTTT TTGAAACTGC
                     CCAAACtGAG TTTCAAAGAT GCGGCGGTCA ACCGCGTGAT GAAACAGATG
                651
                701
                     GCGCCTGCgA TTTTgGGCGT GAgCGTGGCG CAGGTTTCTT TGGTGATCAA
30
                751
                     CACGATTITC GCGTCTTATC TGCAATCGGG CAGCGTTTCA TGGATGTATT
                     ACGCCGACCG CATGATGGAG CTGCCCAGCG GCGTGCTGGG GGCGGCACTC
                801
                     GGTACGATTT TGCTGCCGAC TTTGTCCAAA CACTCGGCAA ACCAAGATAC
                851
                901
                     GGAACAGTTT TCCGCCCTGC TCGACTGGGG TTTGCGCCTG TGCATGCtgc
                     TGACGCTGCC GGCGgcGGTC GGACTGGCGG TGTTGTCGTT cCCgCtGGTG
                951
               1001
35
                     GCGACGCTGT TTATGTACCG CGWATTTACG CTGTTTGACG CGCAGATGAC
                     GCAACACGCG CTGATTGCCT ATTCTTTCGG TTTAATCGGC TTAATCATGA
               1051
               1101
                     TTAAAGTGTT GGCACCCGGC TTCTATGCGC GGCAAAACAT CAAWAMGCCC
                     GTCAAAATCG CCATCTTCAC GCTCATCTGC mCGCAGTTGA TGAACCTTGs
               1151
                     CTTTAYCGGC CCACTrrAAC rCastCGGAC TTTCGCTTGC CATCGGTCTG
               1201
40
                     GGCGCGTGTA TCAATGCCGG ATTGTTGTTT TACCTGTTGC GCAGACACGG
               1251
                     TATTTACCAA CCTGG.CAAG GGTTGGGCAG CGTTCTT.AG CAAAAATGCT
               1301
               1351
                     GCTCTCGCTC GCCGTGA
```

This corresponds to the amino acid sequence <SEQ ID 114; ORF20>:

```
1 MNMLGALAKV GSLTMVSRVL GFVRDTVIAR AFGAGMATDA FFVAFKLPNL
45
                 51
                    LRRVFAEGAF AQAFVPILAE YKETRSKEAX EAFIRHVAGM LSFVLVIVTA
                     LGILAAPWVI YVSAPSFAQD ADKFQLSIDL LRITFPYILL ISLSSFVGSV
                    LNSYHKFGIP AFTPXFLNVS FIVFALFFVP YFDPPVTAXA WAVFVGGILQ
                151
                     LXFQLPWLAK LGFLKLPKLS FKDAAVNRVM KQMAPAILGV SVAQVSLVIN
                251
                     TIFASYLQSG SVSWMYYADR MMELPSGVLG AALGTILLPT LSKHSANODT
50
                     EQFSALLDWG LRLCMLLTLP AAVGLAVLSF PLVATLFMYR XFTLFDAQMT
                301
                351
                     QHALIAYSFG LIGLIMIKVL APGFYARQNI XXPVKIAIFT LICXQLMNLX
                     FXGPLXXIGL SLAIGLGACI NAGLLFYLLR RHGIYQPXQG LGSVLXQKCC
                401
                451
                     SRSP*
```

These sequences were elaborated, and the complete DNA sequence <SEQ ID 115> is:

<sup>55 1</sup> ATGAATATGC TGGGAGCTTT GGCAAAAGTC GGCAGCCTGA CGATGGTGTC
51 GCGCGTTTTG GGATTTGTGC GCGATACGGT CATTGCGCGG GCATTCGGCG

```
CGGGTATGGC GACGGATGCG TTTTTTGTCG CGTTCAAACT GCCCAACCTG
                101
                151
                     CTTCGCCGCG TGTTTGCGGA GGGGGCGTTT GCCCAAGCGT TTGTGCCGAT
                201
                     TTTGGCGGAA TACAAGGAAA CGCGTTCAAA AGAGGCGGCG GAGGCTTTTA
                     TCCGCCATGT GGCGGGGATG CTGTCGTTTG TACTGGTTAT CGTTACCGCG
                251
 5
                     CTGGGCATAC TTGCCGCGCC TTGGGTGATT TATGTTTCCG CACCCGGTTT
                301
                     TGCCCAAGAT GCCGACAAAT TTCAGCTCTC CATCGATTTG CTGCGGATTA
                351
                     CGTTTCCTTA TATATTATTG ATTTCCCTGT CTTCATTTGT CGGCTCGGTA
                401
                451
                     CTCAATTCTT ATCATAAGTT CGGCATTCCG GCGTTTACGC CCACGTTTCT
                     GAACGTGTCG TTTATCGTAT TCGCGCTGTT TTTCGTGCCG TATTTCGATC
                501
10
                     CGCCCGTTAC CGCGCTGGCG TGGGCGGTCT TTGTCGGCGG CATTTTGCAA
                551
                     CTCGGCTTCC AACTGCCCTG GCTGGCGAAA CTGGGCTTTT TGAAACTGCC
                601
                     CAAACTGAGT TTCAAAGATG CGGCGGTCAA CCGCGTGATG AAACAGATGG
                651
                701
                     CGCCTGCGAT TTTGGGCGTG AGCGTGGCGC AGGTTTCTTT GGTGATCAAC
                     ACGATTTTCG CGTCTTATCT GCAATCGGGC AGCGTTTCAT GGATGTATTA
                751
15
                801
                     CGCCGACCGC ATGATGGAGC TGCCCAGCGG CGTGCTGGGG GCGGCACTCG
                851
                     GTACGATTTT GCTGCCGACT TTGTCCAAAC ACTCGGCAAA CCAAGATACG
                901
                     GAACAGTTTT CCGCCCTGCT CGACTGGGGT TTGCGCCTGT GCATGCTGCT
                     GACGCTGCCG GCGGCGGTCG GACTGGCGGT GTTGTCGTTC CCGCTGGTGG
                951
               1001
                     CGACGCTGTT TATGTACCGC GAATTTACGC TGTTTGACGC GCAGATGACG
20
                     CAACACGCGC TGATTGCCTA TTCTTTCGGT TTAATCGGCT TAATCATGAT
               1051
               1101
                     TAAAGTGTTG GCACCCGGCT TCTATGCGCG GCAAAACATC AAAACGCCCG
                     TCAAAATCGC CATCTTCACG CTCATCTGCA CGCAGTTGAT GAACCTTGCC
               1151
                     TTTATCGGCC CACTGAAACA CGTCGGACTT TCGCTTGCCA TCGGTCTGGG
               1201
                     CGCGTGTATC AATGCCGGAT TGTTGTTTTA CCTGTTGCGC AGACACGGTA
               1251
25
               1301
                     TTTACCAACC TGGCAAGGGT TGGGCAGCGT TCTTAGCAAA AATGCTGCTC
                     TCGCTCGCCG TGATGTGCGG CGGACTGTGG GCAGCGCAGG CTTACCTGCC
               1351
               1401
                     GTTTGAATGG GCGCACGCCG GCGGAATGCG GAAAGCGGGG CAGCTCTGCA
               1451
                     TCCTGATTGC CGTCGGCGGC GGACTGTATT TCGCATCACT GGCGGCTTTG
               1501
                     GGCTTCCGTC CGCGCCATTT CAAACGCGTG GAAAACTGA
```

30 This corresponds to the amino acid sequence <SEQ ID 116; ORF20-1>:

```
MNMLGALAKV GSLTMVSRVL GFVRDTVIAR AFGAGMATDA FFVAFKLPNL
                     LRRVFAEGAF AQAFVPILAE YKETRSKEAA EAFIRHVAGM LSFVLVIVTA
                 51
                     LGILAAPWVI YVSAPGFAQD ADKFQLSIDL LRITFPYILL ISLSSFVGSV
                101
                151
                     LNSYHKFGIP AFTPTFLNVS FIVFALFFVP YFDPPVTALA WAVFVGGILO
35
                201
                     LGFQLPWLAK LGFLKLPKLS FKDAAVNRVM KQMAPAILGV SVAQVSLVIN
                251
                     TIFASYLOSG SVSWMYYADR MMELPSGVLG AALGTILLPT LSKHSANODT
                301
                     EQFSALLDWG LRLCMLLTLP AAVGLAVLSF PLVATLFMYR EFTLFDAQMT
                     QHALIAYSFG LIGLIMIKVL APGFYARQNI KTPVKIAIFT LICTQLMNLA
                351
                     FIGPLKHVGL SLAIGLGACI NAGLLFYLLR RHGIYQPGKG WAAFLAKMLL
                401
40
                451
                     SLAVMCGGLW AAQAYLPFEW AHAGGMRKAG QLCILIAVGG GLYFASLAAL
                501
                     GFRPRHFKRV EN*
```

Computer analysis of this amino acid sequence gave the following results:

Homology with the MviN virulence factor of S. typhimurium (accession number P37169) ORF20 and MviN proteins show 63% aa identity in 440aa overlap:

45	Orf20	1	MNMLGALAKVGSLTMVSRVLGFVRDTVIARAFGAGMATDAFFVAFKLPNLLRRVFAEGAF	60
	MviN	14	MN+L +LA V S+TM SRVLGF RD ++AR FGAGMATDAFFVAFKLPNLLRR+FAEGAF MNLLKSLAAVSSMTMFSRVLGFARDAIVARIFGAGMATDAFFVAFKLPNLLRRIFAEGAF	73
50	Orf20	61	AQAFVPILAEYKETRSKEAXEAFIRHVAGMLSFVLVIVTALGILAAPWVIYVSAPSFAQD +QAFVPILAEYK + +EA F+ +V+G+L+ L +VT G+LAAPWVI V+AP FA	120
30	MviN	74	SQAFVPILAEYKSKQGEEATRIFVAYVSGLLTLALAVVTVAGMLAAPWVIMVTAPGFADT	133
	Orf20	121	ADKFQLSIDLLRITFPYILLISLSSFVGSVLNSYHKFGIPAFTPXFLNVSFIVFALFFVP	180
55	MviN	134	ADKF L+ LLRITFPYILLISL+S VG++LN++++F IPAF P FLN+S I FALF P ADKFALTTQLLRITFPYILLISLASLVGAILNTWNRFSIPAFAPTFLNISMIGFALFAAP	193
	Orf20	181	YFDPPVTAXAWAVFVGGILQLXFQLPWLAKLGFLKLPKLSFKDAAVNRVMKQMAPAILGV	240
60	MviN	194	YF+PPV A AWAV VGG+LQL +QLP+L K+G L LP+++F+D RV+KQM PAILGV YFNPPVLALAWAVTVGGVLQLVYQLPYLKKIGMLVLPRINFRDTGAMRVVKQMGPAILGV	253
UU	Orf20	241		300
	MviN	254	SV+Q+SL+INTIFAS+L SGSVSWMYYADR+ME PSGVLG ALGTILLP+LSK A+ + SVSQISLIINTIFASFLASGSVSWMYYADRLMEFPSGVLGVALGTILLPSLSKSFASGNH	313

	Orf20	301 EQFSALLDWGLRLCMLLTLPAAVGLAVLSFPLVATLFMYRXFTLFDAQMTQHALIAYSFG 360
	MviN	+++ L+DWGLRLC LL LP+AV L +L+ PL +LF Y FT FDA MTQ ALIAYS G 314 DEYCRIMDWGLRLCFLLALPSAVALGILAKPLTVSLFQYGKFTAFDAAMTQRALIAYSVG 373
5	Orf20	361 LIGLIMIKVLAPGFYARQNIXXPVKIAIFTLICXQLMNLXFXXXXXXXXXXXXXXXXXXXXXX 420 LIGLI++KVLAPGFY+RQ+I PVKIAI TLI QLMNL F
	MviN	374 LIGLIVVKVLAPGFYSRQDIKTPVKIAIVTLIMTQLMNLAFIGPLKHAGLSLSIGLAACL 433
10	Orf20	421 NAGLLFYLLRRHGIYQPXQG 440
10	MviN	NA LL++ LR+ I+ P G 434 NASLLYWOLRKONIFTPOPG 453

# Homology with a predicted ORF from N. meningitidis (strain A)

ORF20 shows 93.5% identity over a 447aa overlap with an ORF (ORF20a) from strain A of N.

### 15 meningitidis:

		10	20	30	40	50	60
	orf20.pep	MNMLGALAKVGSLTM	VSRVLGFV	RDTVIARAFGA	\GMATDAFFV <i>I</i>	AFKI.PNI.I.PD	FARCAR
	orf20a					[]][]]	
20	OIIZUa	MNMLGALVKVGSLTM	20	RDTVIARAFGA 30			
20	•	10	20	30	40	50	60
		70	80	90	100	110	120
	orf20.pep	AQAFVPILAEYKETR	SKEAXEAF]	RHVAGMLSF	/LVIVTALGII	.AADWUTVUCI	מסמשפט
25	orf20a	- 111111111111111	1111:111				1.11.1
23	OIIZUa	AQAFVPILAEYKETR 70	SKEATEAF) 80	RHVAGMLSFV 90			
		,,	00	90	100	110	120
		. 130	140	150	160	170	180
20	orf20.pep	ADKFQLSIDLLRITE	PYILLISLS	SSFVGSVLNSY	HKFGIPAFTE	YFT.NVCFTVE	מוריים זמי
30	orf20a	111111111111111	5   1   1   1   1   1		111:11111		11111
	OLIZUA	ADKFQLSIDLLRITE	140	SEVGSVLNSY 150	HKFSIPAFTE 160		
		130	140	130	160	170	180
		190	200	210	220	230	240
35	orf20.pep	YFDPPVTAXAWAVFV	GGILOLXFO	)LPWLAKLGFI	KLPKLSFKDA	AUNIDUMEOME	DATICU
	orf20a	# 1   1		1111111111	1111111111	4 1 1 1 1 1 1 1 <del>1 1</del>	11111
	OIIZOA	YFDPPVTALAWAVFV	200	LPWLAKLGFI 210	KLPKLSFKDA 220		
	•	250	200	210	220	230	240
40		250	260	270	280	290	300
	orf20.pep	SVAQVSLVINTIFAS	YLQSGSVSW	MYYADRMMEL	PSGVLGAALG	TILLPTLSKH	CANODE
	orf20a		]	1111111111	1:1111111	1111111111	THE LET I
	OTTZUA	SVAQISLVINTIFAS	11QSGSVSW 260	MIYADRMMEL 270	PGGVLGAALG 280		
45			200	270	280	290	300
		310	320	330	340	350	360
	orf20.pep	EQFSALLDWGLRLCM	LLTLPAAVG	LAVLS FPLVA	TLFMYRXFTL	FDAQMTQHAL	TAVORO
	orf20a	FORSALI DUCI DVCM		1111111:	<u> </u>		111111
50	Olizoa	EQFSALLDWGLRXCM	320	MAVLSEPLVA 330			
		510	320	330	340	350	360
		370	380	390	400	410	420
	orf20.pep	LIGLIMIKVLAPGFY	ARQNIXXPV	KIAIFTLICX	<b>QLMNLXFX</b> GP	LXXIGLSLAI	GLGACI
55	orf20a		11111 : 11	111111111	11111111	1 • 1 1 1 1 1 1	TITI
33	OIIZOa	LIGLIMIKVLAPGFYA	380 AKÖNIKIPV	XIAIFTLICT	QLMNLAFIGP 400		
		. 3.0	300	330	400	410	420
		430	440	450			
60	orf20.pep	NAGLLFYLLRRHGIY(	QPXQGLGSV	LXQKCCSRSP	x		
60	orf20a	NACH EVI I PRICTY	:  ::	:			
	OTTZUG	NAGLLFYLLRRHGIY(	⊋PGKGWA <u>AF</u> 440	LAKMLLSLAV 450			
		450	440	430	460	470	480

The complete length ORF20a nucleotide sequence <SEQ ID 117> is:

<sup>65</sup> ATGAATATGC TGGGAGCTTT GGTAAAAGTC GGCAGCCTGA CGATGGTGTC
51 GCGCGTTTTG GGATTTGTGC GCGATACGGT CATTGCGCGC GCATTCGGCG

<sup>101</sup> CAGGCATGGC GACGGATGCG TTCTTTGTCG CGTTCAAACT GCCCAACCTG

	151 CT	TCGCCGCG	TGTTTGCGGA	GGGGGCGTTT	GCCCAAGCGT	TTGTGCCG	<b>\</b> Τ	
	201 TT	TGGCGGAA	TATAAGGAAA	CGCGTTCTAA	AGAGGCGACG	GAGGCTTTT	î A	
	251 TC 301 CT	CCCCATGT	TTCCCCCCCCC	TTGCCTCATT	TACTGGTCAT TATGTTTCCG	CACCCCC	CG	
5	351 TG	CCAAAGAT	GCCGACAAAT	TTCAGCTCTC	TATCGATTTG	CTGCGGATI	. I	
•	401 CG	TTTCCTTA	TATCTTATTG	ATTTCACTTT	CCTCTTTTGT	CGGCTCGGT	ra.	
	451 CT	CAATTCCT	ATCATAAATT	CAGCATTCCT	GCGTTTACGC	CCACGTTC	CT	
					TTTCGTGCCG			
10					TTGTCGGCGG			
10					CTGGGTTTTT			
					CCGCGTGATG AGATTTCTTT			
					AGCGTTTCAT			
	801 CG	CCGACCGC	ATGATGGAAC	TGCCCGGCGG	CGTGCTGGGG	GCGGCACTC	iG	
15	851 GT	ACGATTTT	GCTGCCGACT	TTGTCCAAAC	ACTCGGCAAA	CCAAGATAC	,c :G	
	901 GA	ACAGTTTT	CCGCCCTGCT	CGACTGGGGT	TTGCGCNTGT	GCATGCTGC	CT	
	951 GA	CGCTGCCG	GCGGCGGTCG	GAATGGCGGT	GTTGTCGTTC	CCGCTGGT	<b>;</b> G	
	1001 CA	ACCTTGTT	TATGTACCGA	GAATTCACGC	TGTTTGACGC	GCAGATGAC	:G	
•	1051 CA	ACACGCGC	TGATTGCCTA	TTCTTTCGGT	TTAATCGGTT	TAATCATG	T	
20	1101 TA	AAGTGTTG	GCGCCCGGCT	TTTATGCGCG	GCAAAACATC	AAAACGCCC	;G	
	1151 TC	AAAATCGC	CATCTTCACG	CTCATTTGCA	CGCAGTTGAT	GAACCTTGC	;c	
	1201 TT 1251 CG	CCTCTATC	AATCCCCCAT	TOTT CONTROL TO	TCGCTTGCCA CCTGTTGCGC	TCGGTCTGG	∍G	
	1301 TT	TACCAACC	TGCCDAGGT	TGGGCAGCGT	TCTTGGCAAA	AGACACGGT	:A	
25	1351 TC	CCTCGCCG	TGATGGGAGG	CGGCCTGTAT	GCCGCCCAAA	TCTGGCTGC	.C *C	
20	1401 GT	TCGACTGG	GCACACGCCG	GCGGAATGCA	AAAGGCCGCC	CGGCTCTTC	ZA.	
	1451 TC	CTGATTGC	CGTCGGCGGC	GGACTGTATT	TCGCATCACT	GGCGGCTTT	ig '	
			CGCGCCATTT					
	771 '				0E0 ID 110	_		•
	This encodes a prot	em naving	g amino acid	sequence <	SEQ ID 118	>:		
30	1 MN	MLGALVKV	GSLTMVSRVL	GFVRDTVIAR	AFGAGMATDA	FFVAFKLPN	1L	
	51 LR	RVFAEGAF	AQAFVPILAE	YKETRSKEAT	EAFIRHVAGM	LSFVLVIVI	'A	
					LRITFPYILL			
					YFDPPVTALA			
35	201 <u>LG</u> 251 TI	FOTEMTAK	LGFLKLPKLS	FKDAAVNRVM	KOMAPAILGV	SVAQISLVI	N	
33	301 EO	PSQLIDGG	TOVOMITADE	MARTINGOATO	AALGTILLPT PLVATLFMYR	LSKHSANQL	)T	
	351 QH	ALTAYSEG	I.TGI.TMTKVI.	APGFYARONT	KTPVKIAIFT	T.TCTOT.MNI	.v	
					RHGIYQPGKG			
		AVMGGGLY	AAQIWLPFDW	AHAGGMQKAA	RLFILIAVGG	GLYFASLA	還	
40	501 GF	RPRHFKRV	ES*					
	ORF20a and ORF2	0-1 show	96.5% ident	ity in 512 aa	overlap:			
			10	20	30	10	50	60
	orf20a.pep	MNMLGA	LVKVGSLTMVS		IARAFGAGMATI			
		111111	1:111111111		[[]]		1111111	1111
45	orf20-1	MNMLGA			LARAFGAGMATI			EGAF
			10	20	30	10	50	60
			70	80		00 1	10	120
	orf20a.pep	AQAFVP	ILAEYKETRSE	<b>CEATEAFIRHV</b>	AGMLSFVLVIV	TALGILAAPW	VIYVSAPG	FAKD
50		111111	11111111111		14111111111			11:1
	orf20-1	AQAFVP			AGMLSFVLVIV			
			70	80	90 10	00 1	110	120
			130				.70	180
55	orf20a.pep	ADKFQL	SIDLLRITFPY	(ILLISLSSFV	SSVLNSYHKFS:	PAFTPTFLN	<b>VSFIVFAL</b>	FFVP
		111111		шинин	H		111111111	
	orf20-1	ADKFQL			SSVLNSYHKFG:			
			130	140	L50 10	50 1	170	180
60			190	200	210 22	20 2	230	240
00	orf20a.pep	YFDDDU			LAKLGFLKLPK			240
	orrrow.bob	111111	11111111111				.venQeAPA;	1111
	orf20-1	YFDPPV	TALAWAVEVG	SILQLGFOLPW	LAKLGFLKLPK	SFKDAAVNE	VMKOMAPA.	ILGV
<u>.</u>			190				30	240
65								-
		0	250				290	300
	orf20a.pep	SVAQIS	LVINTIFASYI	LQSGSVSWMYY	ADRMMELPGGV	LGAALGTILI	PTLSKHSA	TOON

orf20-1						
	250	260	270	280	290	300
	310	320	330	340	350	. 360
orf20a.pep	EQFSALLDWGLRXCM			TLFMYREFTL	FDAOMTOHAL	
orf20-1	EOFSALLOWGIRLOM					TAVEEC
01120 1			330			360
	370	380	390	400	410	420
orf20a.pep	LIGLIMIKVLAPGFY	ARQNIKTPVK	IAIFTLICT(	<b>QLMNLAFIGP</b>	LKHVGLSLAI	GLGACI
orf20-1	LIGLIMIKVLAPGFY	 ARQNIKTPVK	 IAIFTLICT(			 GLGACI
	370	380	390	400	410	420
•	420	440	450	460	470	
orf20a.pep						480 (CMOKAA
0000000						
orf20-1						
	430	440	450	460	470	480
	490	500	510			
orf20a.pep				•		
012204.505	:1 11111111111		111111:1			
orf20-1	QLCILIAVGGGLYFA	SLAALGERPR	HFKRVENX			
	490	500	510			
	orf20a.pep orf20a.pep orf20a.pep orf20-1 orf20a.pep orf20-1	orf20-1         SVAQVSLVINTIFAS 250           310         310           orf20a.pep         EQFSALLDWGLRXCM	orf20-1         SVAQVSLVINTIFASYLQSGSVSWM           250         260           310         320           orf20a.pep         EQFSALLDWGLRXCMLLTLPAAVGM                                     orf20-1         EQFSALLDWGLRLCMLLTLPAAVGL           310         320           370         380           orf20a.pep         LIGLIMIKVLAPGFYARQNIKTPVK           370         380           orf20-1         LIGLIMIKVLAPGFYARQNIKTPVK           370         380           orf20a.pep         NAGLLFYLLRHGIYQPGKGWAAFI  orf20a.pep         NAGLLFYLLRHGIYQPGKGWAAFI           430         440           490         500           orf20a.pep         RLFILIAVGGGLYFASLAALGFRPF           :	orf20-1         SVAQVSLVINTIFASYLQSGSVSWMYYADRMMELT           250         260         270           310         320         330           orf20a.pep         EQFSALLDWGLRXCMLLTLPAAVGMAVLSFPLVAT           iiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiii	orf20-1         SVAQVSLVINTIFASYLQSGSVSWMYYADRMMELPSGVLGAALG           250         260         270         280           310         320         330         340           orf20a.pep         EQFSALLDWGLRXCMLLTLPAAVGMAVLSFPLVATLFMYREFTL	orf20-1         SVAQVSLVINTIFASYLQSGSVSWMYYADRMMELPSGVLGAALGTILLPTLSKH           250         260         270         280         290           310         320         330         340         350           orf20a.pep         EQFSALLDWGLRXCMLLTLPAAVGMAVLSFPLVATLFMYREFTLFDAQMTQHAL 

# Homology with a predicted ORF from N.gonorrhoeae

30 ORF20 shows 92.1% identity over a 454aa overlap with a predicted ORF (ORF20ng) from N. gonorrhoeae:

	orf20.pep	MNMLGALAKVGSLTMVSRVLGFVRDTVIARAFGAGMATDAFFVAFKLPNLLRRVFAEGAF	60
35	orf20ng		60
33	orf20.pep	AQAFVPILAEYKETRSKEAXEAFIRHVAGMLSFVLVIVTALGILAAPWVIYVSAPSFAQD	120
	orf20ng	AQAFVPILAEYKETRSKEATEAFIRHVAGMLSFVLIVVTALGILAAPWVIYVSAPGFTKD	120
40	orf20.pep	ADKFQLSIDLLRITFPYILLISLSSFVGSVLNSYHKFGIPAFTPXFLNVSFIVFALFFVP	180
	orf20ng	ADKFQLSISLLRITFPYILLISLSSFVGSILNSYHKFGIPAFTPTFLNISFIVFALFFVP	180
45	orf20.pep	YFDPPVTAXAWAVFVGGILQLXFQLPWLAKLGFLKLPKLSFKDAAVNRVMKQMAPAILGV	240
	orf20ng	YFDPPVTALAWAVFVGGILQLGFQLPWLAKLGFLKLPKLNFKDAAVNRVMKQMAPAILGV	240
	orf20.pep	SVAQVSLVINTIFASYLQSGSVSWMYYADRMMELPSGVLGAALGTILLPTLSKHSANQDT	300
50	orf20ng	SVAQISLVINTIFASYLQSGSVSWMYYADRMMELPGGVLGAALGTILLPTLSKHSANQDT	300
	orf20.pep	EQFSALLDWGLRLCMLLTLPAAVGLAVLSFPLVATLFMYRXFTLFDAQMTQHALIAYSFG	360
55	orf20ng	EQFSALLDWGLRLCMLLTLPAAAGLAVLSFPLVATLFMYREFTLFDAQMTQHALIAYSFG	360
33	orf20.pep	LIGLIMIKVLAPGFYARQNIXXPVKIAIFTLICXQLMNLXFXGPLXXIGLSLAIGLGACI	420
	orf20ng	LIGLIMIKVLASGFYARQNIKTPVKIAIFTLICTQLMNLAFIGPLKHAGLSLAIGLGACI	420
60	orf20.pep	NAGLLFYLLRRHGIYQPXQGLGSVLXQKCCSRSP 454	
	orf20ng	NAGLLFFLFRKHGIYRPGQGLGQPSWRKCCSRSP 454	

An ORF20ng nucleotide sequence <SEQ ID 119> was predicted to encode a protein having amino acid sequence <SEQ ID 120>:

```
MNMLGALAKV GSLTMVSRVL GFVRDTVIAR AFGAGMATDA FFVAFKLPNL
                    LRRVFAEGAF AQAFVPILAE YKETRSKEAT EAFIRHVAGM LSFVLIVVTA
                51
                    LGILAAPWVI YVSAPGFTKD ADKFQLSISL LRITFPYILL ISLSSFVGSI
               101
                    LNSYHKFGIP AFTPTFLNIS FIVFALFFVP YFDPPVTALA WAVFVGGILQ
               151
 5
                    LGFQLPWLAK LGFLKLPKLN FKDAAVNRVM KQMAPAILGV SVAQISLVIN
               201
                    TIFASYLOSG SVSWMYYADR MMELPGGVLG AALGTILLPT LSKHSANQDT
               251
                    EQFSALLDWG LRLCMLTLP AAAGLAVLSF PLVATLFMYR EFTLFDAQMT
               301
                     QHALIAYSFG LIGLIMIKVL ASGFYARQNI KTPVKIAIFT LICTQLMNLA
               351
                     FIGPLKHAGL SLAIGLGACI NAGLLFFLFR KHGIYRPGQG LGQPSWRKCC
               401
10
               451
                     SRSP*
     Further DNA sequence analysis revealed the following DNA sequence <SEQ ID 121>:
                    ATGAATATGC TTGGAGCTTT GGCAAAAGTC GGCAGCCTGA CGATGGTGTC
                 1
                     GCGCGTTTTG GGATTTGTGC GCGATACGGT CATTGCGCGG GCATTCGGCG
                51
                     CGGGTATGGC GACGGATGCG TTTTTTGTCG CGTTCAAACT GCCCAACCTG
                101
15
                     CTTCGCCGCG TGTTTGCGGA GGGGGCGTTT GCCCAAGCGT TTGTGCCGAT
               151
                     TTTGGCGGAA TATAAGGAAA CGCGTTCTAA AGAGGCGAcg gAGGCTTTTA
                201
                    TCCGCCACGt tgcgggAatg CTGTCGTTTG TGCTGATcgt cGttacCGCG CTGGGCATAC TTGCCGCgcc tTGGGTGATT TATGTTtccg CgcccGGCTT
                251
                301
                    TACCAAAGAC GCGGACAAGT TCCAACTTTC CATCAGCCTG CTGCGGATTA
                351
                     CGTTTCCTTA TATATTATTG ATTTCTTTGT CTTCTTTTGT CGGCTCGATA
20
                401
                     CTCAATTCCT ACCATAAGTT CGGCATTCCC GCGTTTACGC CCACGTTTTT
                451
                     AAACATCTCT TTTATCGTAT TCGCACTGTT TTTCGTGCCG TATTTCGATC
                501
                    CGCCCGTTAC CGCGCTGGCG TGGGCGGTTT TTGTCGGCGG TATTTTGCAG
                551
                    CTCGGTTTCC AACTGCCGTG GCTGGCGAAA CTGGGCTTTT TGAAACTGCC
                601
25
                651
                     CAAACTGAAT TTCAAAGATG CGGCGGTCAA CCGCGTCATG AAACAGATGG
                     CGCCTGCGAT TTTGGGCGTG agcgTGGCGC AAATTTCTTT GgttATCAAC
                701
                751 ACGATTTCG CGTCTTATCT GCAATCGGGC AGCGTTTCAT GGATGTatta
                801
                     cgCCGACCGC ATGATGGAGc tgcgccGGGG CGTGCTGGGG GCTGCACTCG
                851
                     GTACAATTTT GCTGCCGACT TTGTCCAAAC ACTCGGCAAA CCAAGATACG
30
                901
                     GAACAGTTTT CCGCCCTGCT CGACTGGGGT TTGCGCCTGT GCATGCTGCT
                951 GACGCTGCCG GCGGCGGccg GACTGGCGGT ATTGTCGTTC CCGCTGGTGG
               1001
                     CGACGCTGTT TATGTACCGA GAATTCACGC TGTTTGACGC ACAAATGACG
                     CAACACGCGC TGATTGCCTA TTCTTTCGGT TTAATCGGTT TAATTATGAT
               1051
               1101
                     TAAAGTGTTG GCATCCGGCT TTTATGCGCG GCAAAACATC AAAACGCCCG
                     TCAAAATCGC CATCTTCACG CTCATCTGCA CGCAGTTGAT GAACCTCGCC
35
               1151
               1201
                     TTTATCGGTC CGTTGAAACA CGCCGGGCTT TCGCTCGCCA TCGGCCTGGG
                     CGCGTGCATC AACGCCGGAT TGTTGTTCTT CCTGTTGCGC AAACACGGTA
               1251
               1301
                     TTTACCGGCC cggcaggggt tgggcggcgt TCTTGGCGAA AATGCTGCTC
                     GCGCTCGCCG TGATGTGCGG CGGACTGTGG GCGGCGCAGG CTTGCCTGCC
               1351
40
               1401
                     GTTCGAATGG GCGCACGCCG GCGGAATGCG GAAAGCGGGG CAGCTCTGCA
                     TCCTGATTGC CGTCGGCGGC GGACTGTATT TCGCATCTCT GGCGGCTTTG
               1451
                     GGCTTCCGTC CGCGCCATTT CAAACGCGTG GAAAGCTGA
               1501
      This encodes the following amino acid sequence <SEQ ID 122; ORF20ng-1>:
                     MNMLGALAKV GSLTMVSRVL GFVRDTVIAR AFGAGMATDA FFVAFKLPNL
45
                     LRRVFAEGAF AQAFVPILAE YKETRSKEAT EAFIRHVAGM LSFVLIVVTA
                     LGILAAPWVI YVSAPGFTKD ADKFQLSISL LRITFPYILL ISLSSFVGSI
                101
                     LNSYHKFGIP AFTPTFLNIS FIVFALFFVP YFDPPVTALA WAVFVGGILQ
                     LGFQLPWLAK LGFLKLPKLN FKDAAVNRVM KQMAPAILGV SVAQISLVIN
                201
                     TIFASYLQSG SVSWMYYADR MMELRRGVLG AALGTILLPT LSKHSANQDT
                251
 50
                     EQFSALLDWG LRLCMLLTLP AAAGLAVLSF PLVATLFMYR EFTLFDAQMT
                301
                     QHALIAYSFG LIGLIMIKVL ASGFYARQNI KTPVKIAIFT LICTQLMNLA
                351
                     FIGPLKHAGL SLAIGLGACI NAGLLFFLLR KHGIYRPGRG WAAFLAKMLL
                401
                     ALAVMCGGLW AAQACLPFEW AHAGGMRKAG QLCILIAVGG GLYFASLAAL
                451
                501 GFRPRHFKRV ES*
      ORF20ng-1 and ORF20-1 show 95.7% identity in 512 aa overlap:
 55
                                                                40
                                                     30
                                                                          50
                          MNMLGALAKVGSLTMVSRVLGFVRDTVIARAFGAGMATDAFFVAFKLPNLLRRVFAEGAF
           orf20-1.pep
                          orf20ng-1
                          MNMLGALAKVGSLTMVSRVLGFVRDTVIARAFGAGMATDAFFVAFKLPNLLRRVFAEGAF
 60
                                  10
                                                      30
                                                                40
                                                      90
                                                               100
                                                                         110
                          AQAFVPILAEYKETRSKEAAEAFIRHVAGMLSFVLVIVTALGILAAPWVIYVSAPGFAQD
           orf20-1.pep
                          65
                          AOAFVPILAEYKETRSKEATEAFIRHVAGMLSFVLIVVTALGILAAPWVIYVSAPGFTKD
           orf20ng-1
```

			•				
		70	80	90	100	110	120
	orf20-1.pep	130 ADKFQLSIDLLRIT	140	150	160	170	180
5	01120 1.pop						
	orf20ng-1	ADKFQLSISLLRIT					
		130	140	150	160	170	180
10		190 YFDPPVTALAWAVF	200	210	220	230	240
10	orf20-1.pep						
	orf20ng-1	YFDPPVTALAWAVF	VGGILQLGF	QLPWLAKLGFL	KLPKLNFKDA	AAVNRVMKQMI	APAILGV
		190	200	210	220	230	240
15		250	260	270	280	290	300
	orf20-1.pep	SVAQVSLVINTIFA				STILLPTLSKI 	
	orf20ng-1	SVAQISLVINTIFA	SYLOSGSVS	WMYYADRMMEL	RRGVLGAAL	GTILLPTLSKI	ISANODT
20	•	250	260	270	280	290	300
20		310	320	330	340	350	360
	orf20-1.pep	EQFSALLDWGLRLC	MLLTLPAAV	GLAVLSFPLVA	TLFMYREFT]	LFDAQMTQHA]	LIAYSFG
	orf20ng-1	EQFSALLDWGLRLC	MLLTLPAAA	GLAVLSFPLVA	TLFMYREFT	LFDAQMTQHAI	LIAYSFG
25	•	310	320	330	340	350	360
		370	380	390	400	410	420
	orf20-1.pep	LIGLIMIKVLAPGF	YARQNIKTP	VKIAIFTLICT	QLMNLAFIGI	PLKHVGLSLA:	IGLGACI
30	orf20ng-1	LIGLIMIKVLASGE	YARQNIKTP	VKIAIFTLICT	QLMNLAFIG	PLKHAGLSLA:	IGLGACI
		370	380	390	400	410	420
	•	430	440	450	460	470	480
35	orf20-1.pep	NAGLLFYLLRRHGI					
Ÿ.	orf20ng-1	NAGLLFFLLRKHGI	YRPGRGWAA	<b>FLAKMLLALA</b> V	MCGGLWAAQ	ACLPFEWAHA	GMRKAG
		430	440	450	460	470	480
40		490	500	510			•
40	orf20-1.pep	QLCILIAVGGGLYE			•		
	orf20ng-1	QLCILIAVGGGLYE	ASLAALGFR	PRHFKRVESX			
		490	500	510			
	In addition, ORF20n	g-1 shows signifi	cant homo	logy with a	virulence fa	actor of S.ty	phimurium:
45	splP37169 MVIN	SALTY VIRULENC	E FACTOR	MVIN pirlls	540271 mvi	N protein	- Salmonella
	typhimurium o	gi 438252 (Z261	.33) mviB	gene pro	oduct [Sa	lmonella	typhimurium]
		21 (D25292) ORE (750.1 bits), E					
<b>c</b> 0		309/467 (66%),					
50	Query: 1 N	INMLGALAKVGSLTMV	SRVLGFVRD	TVIARAFGAGN	/ATDAFFVAF	KT.PNT.T.RRVF	AFGAF 60
		N+L +LA V S+TM	SRVLGF RD	++AR FGAGN	(ATDAFFVAF	KLPNLLRR+F	AEGAF
	Sbjct: 14 N	MNLLKSLAAVSSMTME	SRVLGFARD	AIVARIFGAGN	(ATDAFFVAF	KLPNLLRRIF	AEGAF 73
55		AQAFVPILAEYKETRS					
		+QAFVPILAEYK + SQAFVPILAEYKSKQ0					
							•
60	Query: 121 1	ADKFQLSISLLRITFE ADKF L+ LLRITFE	PYILLISLSS PYILLISL+S	FVGSILNSYHE VG+ILN+++	KFGIPAFTPT: FF IPAF PT:	FLNISFIVFA FLNIS I FA	LFFVP 180
		ADKFALTTQLLRITF					
	Query: 181	YFDPPVTALAWAVFVO	GILOLGFOL	PWLAKI.GFI.KI	.PKLNFKNAA	UNRUMKOMA D	ATT.CV 240
	•	YF+PPV ALAWAV VO	GG+LQL +QL	P+L K+G L I	LP++NF+D	RV+KQM P	AILGV
65	Sbjct: 194	YFNPPVLALAWAVTV	GVLQLVYQL	PYLKKIGMLVI	LPRINFRDTG	AMRVVKQMGP.	AILGV 253
		SVAQISLVINTIFAS)					
		SV+QISL+INTIFAS: SVSQISLIINTIFAS:			GVLG ALGT	ILLP+LSK	A+ +
70	00,000	o.ogidhiiniiinoi	v one	T TUDITHER E	CARCANIOI	- mut 3 F3V2 [.	MOGNU 313
	•						

```
301 EQFSALLDWGLRLCMLLTLPAAAGLAVLSFPLVATLFMYREFTLFDAQMTQHALIAYSFG 360
          Query:
                        +++ L+DWGLRLC LL LP+A L +L+ PL +LF Y +FT FDA MTQ ALIAYS G
                    314 DEYCRLMDWGLRLCFLLALPSAVALGILAKPLTVSLFQYGKFTAFDAAMTQRALIAYSVG 373
          Sbjct:
                    361 LIGLIMIKVLASGFYARQNIKTPVKIAIFTLICTQLMNLAFIGPLKHAGLSLAIGLGACI 420
5
          Query:
                        LIGLI++KVLA GFY+RQ+IKTPVKIAI TLI TQLMNLAFIGPLKHAGLSL+IGL AC+
                    374 LIGLIVVKVLAPGFYSRQDIKTPVKIAIVTLIMTQLMNLAFIGPLKHAGLSLSIGLAACL 433
          Sbjct:
                    421 NAGLLFFLLRKHGIYRPGRGWXXXXXXXXXXXXVMCGGLWAAQACLP 467
          Query:
                        NA LL++ LRK I+ P GW
10
                                                         VM
                                                             L+
                                                                     +P
                    434 NASLLYWQLRKQNIFTPQPGWMWFLMRLIISVLVMAAVLFGVLHIMP 480
          Sbjct:
           Score = 70 (33.4 bits), Expect = 1.1e-220, Sum P(2) = 1.1e-220
           Identities = 14/41 (34%), Positives = 23/41 (56%)
15
                    469 EWAHAGGMRKAGQLCILIAVGGGLYFASLAALGFRPRHFKR 509
           Query:
                              + + +L ++ G
                                              YFA+LA LGF+ + F R
           Sbjct:
                    481 EWSQGSMLWRLLRLMAVVIAGIAAYFAALAVLGFKVKEFVR 521
```

Based on this analysis, including the homology with a virulence factor from *S.typhimurium*, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 15

The following partial DNA sequence was identified in N. meningitidis <SEQ ID 123>:

```
25
                     atGATTAAAA TCAAAAAAGG TCTAAACCTG CCCATCGCGG GCAGACCGGA
                     GCAAGCCGTT tACGACGGCC CGGCCaTTAC CGAAGtCGCG TTGCTTGGCG
                 51
                    AAGAATATGC CGGTATGCGC CCCTCGATGA AAGTCAAGGA AGGCGATGCC
                101
                151
                     GTCAAAAAAG GCCAAGTGCT GTTTGAAGAC AAAAAGAATC CGGGCGTGGT
                     GTTTACTGCG CCGGCTTCAG GCAAAATCGC CGCGATTCAC CGTGGCGAAA
                201
30
                     AGCGCGTACT TCAGTCAGTC GTGATTGCCG TTGAArGCAA CGACGAAATC
                     GAGTTTGAAC GCTACGCACC TGAAGCGCTG GCAAACTTAA GCGGCGAAGA
                301
                     AGTGCGCCGC AACCTGATCC AATCCGGTTT GTGGACTGCG CTGCGCACCC
                     GTCCGTTCAG CAAAATTCCT GCCGTCGATG CCGAGCCGTT CGCCATCTTC
                401
                    GTCAATGCGA tGGACACCAA TCCG..
```

This corresponds to the amino acid sequence <SEQ ID 124; ORF22>:

```
1 MIKIKKGLNL PIAGRPEQAV YDGPAITEVA LLGEEYAGMR PSMKVKEGDA
51 VKKGQVLFED KKNPGVVFTA PASGKIAAIH RGEKRVLQSV VIAVEXNDEI
101 EFERYAPEAL ANLSGEEVRR NLIQSGLWTA LRTRPFSKIP AVDAEPFAIF
151 VNAMDTNP..
```

40 Further work revealed the complete nucleotide sequence <SEQ ID 125>:

```
1 ATGATTAAAA TCAAAAAAGG TCTAAACCTG CCCATCGCGG GCAGACCGGA
                51
                    GCAAGCCGTT TACGACGGCC CGGCCATTAC CGAAGTCGCG TTGCTTGGCG
                    AAGAATATGC CGGTATGCGC CCCTCGATGA AAGTCAAGGA AGGCGATGCC
               101
               151
                    GTCAAAAAG GCCAAGTGCT GTTTGAAGAC AAAAAGAATC CGGGCGTGGT
45
                    GTTTACTGCG CCGGCTTCAG GCAAAATCGC CGCGATTCAC CGTGGCGAAA
               201
                    AGCGCGTACT TCAGTCAGTC GTGATTGCCG TTGAAGGCAA CGACGAAATC
               251
                    GAGTTTGAAC GCTACGCACC TGAAGCGCTG GCAAACTTAA GCGGCGAAGA
                    AGTGCGCCGC AACCTGATCC AATCCGGTTT GTGGACTGCG CTGCGCACCC
               351
                    GTCCGTTCAG CAAAATTCCT GCCGTCGATG CCGAGCCGTT CGCCATCTTC
                401
50
                451
                    GTCAATGCGA TGGACACCAA TCCGCTGGCT GCCGACCCTA CGGTCATTAT
                    CAAAGAAGCC GCCGAGGATT TCAAACGCGG CCTGTTGGTA TTGAGCCGTT
               501
                    TGACCGAACG CAAAATCCAT GTTTGTAAGG CAGCTGGCGC AGACGTGCCG
                551
                     TCTGAAAATG CTGCCAACAT CGAAACACAT GAATTCGGCG GCCCGCATCC
                    TGCCGGTTTG AGTGGCACGC ACATTCATTT CATCGAGCCG GTCGGCGCGA
                651
55
                    ATAAAACCGT GTGGACCATC AATTATCAAG ATGTAATTAC CATTGGCCGT
                    TTGTTTGCAA CAGGCCGTCT GAACACCGAG CGCGTGATTG CCCTAGGTGG
                751
                     TTCTCAAGTC AACAAACCGC GCCTCTTGCG TACCGTTTTG GGTGCGAAAG
                801
                     TATCGCAAAT TACTGCGGGC GAATTGGTTG ACACAGACAA CCGCGTGATT
                851
                    TCCGGTTCGG TATTGAACGG CGCGATTACA CAAGGCGCGC ACGATTATTT
```

orf22a

```
GGGACGCTAC CACAATCAGA TTTCCGTTAT CGAAGAAGGC CGCAGCAAAG
               951
                     AGCTGTTCGG CTGGGTTGCG CCGCAGCCGG ACAAATACTC CATCACGCGT
              1001
                     ACAACCCTCG GCCATTTCCT GAAAAACAAA CTCTTCAAGT TCAACACAGC
              1051
                     CGTCAACGGC GGCGACCGCG CCATGGTGCC GATTGGTACT TACGAGCGCG
              1101
5
              1151
                     TGATGCCCTT GGATATCCTG CCCACCCTGC TTTTGCGCGA TTTAATCGTC
                     GGCGATACCG ACAGCGCGCA GGCATTGGGT TGCTTGGAAT TGGACGAAGA
              1201
              1251
                     AGACCTCGCT TTGTGCAGCT TCGTCTGCCC GGGCAAATAC GAATACGGCC
              1301 CGCTGTTGCG CAAAGTGCTG GAAACCATTG AGAAGGAAGG CTGA
     This corresponds to the amino acid sequence <SEQ ID 126; ORF22-1>:
10
                    MIKIKKGLNL PIAGRPEQAV YDGPAITEVA LLGEEYAGMR PSMKVKEGDA
VKKGQVLFED KKNPGVVFTA PASGKIAAIH RGEKRVLQSV VIAVEGNDEI
                 51
                     EFERYAPEAL ANLSGEEVRR NLIQSGLWTA LRTRPFSKIP AVDAEPFAIF
                101
                     VNAMDTNPLA ADPTVIIKEA AEDFKRGLLV LSRLTERKIH VCKAAGADVP
                201
                     SENAANIETH EFGGPHPAGL SGTHIHFIEP VGANKTVWTI NYQDVITIGR
15
                     LFATGRINTE RVIALGGSQV NKPRLIRTVL GAKVSQITAG ELVDTDNRVI
                251
                     SGSVLNGAIT QGAHDYLGRY HNQISVIEEG RSKELFGWVA PQPDKYSITR
                301
                351
                    TTLGHFLKNK LFKFNTAVNG GDRAMVPIGT YERVMPLDIL PTLLLRDLIV
                401 GDTDSAQALG CLELDEEDLA LCSFVCPGKY EYGPLLRKVL ETIEKEG*
     Further work identified the corresponding gene in strain A of N. meningitidis <SEQ ID 127>:
20
                  1 ATGATTAAAA TCAAAAAAGG TCTAAACCTG CCCATCGCGG GCAGACCGGA
                 51
                     GCAAGTCATT TATGACGGGC CCGTCATTAC CGAAGTCGCG TTGCTTGGCG
                     AAGAATATGC CGGTATGCGC CCCTNGATGA AAGTCAAGGA AGGCGATGCC
                101
                     GTCAAAAAAG GCCAAGTGCT GTTTGAAGAC AAAAAGNATC CGGGCGTGGT
                151
                201
                     GTTTACCGCG CCNGTTTCAG GCAAAATCGC CGCCATCCAT CGCGGCGAAA
25
                251 AGCGCGTACT TCAGTCGGTC GTGATTGCCG TTGAAGGCAA CGACGAAATC
                301
                     GAGTTCGAAC GCTACGCGCC CGAAGCGTTG GCAAACTTAA GCGGCGANGA
                351 ANTHNGNNGC AATCTGATCC AATCCGGTTT GTGGACTGCG CTGCGTANCC
                     GTCCGTTCAG CAAAATCCCT GCCGTCGATG CCGAGCCGTT CGCCATCTTC
                401
                451
                     GTCAATGCGA TGGACACCAA TCCGCTNGCG GCAGACCCTG TGGTTGTGAT
30
                     CAAAGAAGCC GNCGANGATT TCAGACGANG TNTGCTGGTA TTGAGCCGTT
                501
                     TGACCGAGCG TAAAATCCAT GTGTGTAAGG CAGCTGGCGC AGACGTGCCG
                551
                     TCTGAAAATG CTGCCAACAT CGAAACACAT GAATTCGGCG GCCCGCATCC
                601
                651
                     GGCCGGTTTG AGTGGCACGC ACATTCATTT CATTGAGCCG GTCGGTGCAA
                     ACAAAACCGT TTGGACCATC AATTATCAAG ATGTAATTGC CATCGGACGT
                701
                     TTGTTTGCAA CAGGCCGTCT GAACACCGAG CGCGTGATTG CTTTGGGTGG
35
                751
                     TTCTCAAGTC AACAAACCAC GCCTCTTGCG TACCGTTTTG GGTGCGAAAG
                801
                851
                     TATCGCAAAT TACTGCGGGC GAATTGGTTG ACGCAGACAA CCGCGTGATT
                     TCCGGTTCGG TATTGAACGG CGCGATTACA CAAGGCGCGC ACGATTATTT
                901
                     GGGACGCTAC CACAATCAGA TTTCCGTTAT CGAAGAAGGC CGCAGCAAAG
                951
40
                     AGCTGTTCGG CTGGGTTGCG CCGCAGCCGG ACAAATACTC CATCACGCGT
               1001
               1051
                     ACGACCCTCG GCCATTTCCT GAAAAACAAA CTCTTCAAGT TCACGACAGC
                     CGTCAACGGT GGCGACCGCG CCATGGTGCC GATTGGTACT TACGAGCGCG
               1101
               1151
                     TAATGCCGCT AGACATCCTG CCTACCCTGC TTTTGCGCGA TTTAATCGTC
                     GGCGATACCG ACAGCGCGCA AGCATTGGGT TGCTTGGAAT TGGACGAAGA
               1201
45
                     AGACCTCGCT TTGTGCAGCT TCGTCTGCCC GGGCAAATAC GAATANGGCC
               1251
                     CGCTGTTGCG TAAGGTGCTG GAAACCNTTG AGAAGGAAGG CTGA
               1301
      This encodes a protein having amino acid sequence <SEQ ID 128; ORF22a>:
                  1 MIKIKKGLNL PIAGRPEQVI YDGPVITEVA LLGEEYAGMR PXMKVKEGDA
                     VKKGQVLFED KKXPGVVFTA PVSGKIAAIH RGEKRVLQSV VIAVEGNDEI
50
                101 EFERYAPEAL ANLSGXEXXX NLIQSGLWTA LRXRPFSKIP AVDAEPFAIF
                151 VNAMDTNPLA ADPVVVIKEA XXDFRRXXLV LSRLTERKIH VCKAAGADVP
                     SENAANIETH EFGGPHPAGL SGTHIHFIEP VGANKTVWTI NYQDVIAIGR
                201
                251
                     LFATGRLNTE RVIALGGSQV NKPRLLRTVL GAKVSQITAG ELVDADNRVI
                301
                     SGSVLNGAIT QGAHDYLGRY HNQISVIEEG RSKELFGWVA PQPDKYSITR
55
                     TTLGHFLKNK LFKFTTAVNG GDRAMVPIGT YERVMPLDIL PTLLLRDLIV
                351
                401 GDTDSAQALG CLELDEEDLA LCSFVCPGKY EXGPLLRKVL ETXEKEG*
      The originally-identified partial strain B sequence (ORF22) shows 94.2% identity over a 158aa
      overlap with ORF22a:
                                10
                                           20
                                                     30
                                                               40
                        MIKIKKGLNLPIAGRPEQAVYDGPAITEVALLGEEYAGMRPSMKVKEGDAVKKGQVLFED
60
           orf22.pep
```

 -124-

		10	20	30	40	50	60	
	orf22.pep	70 KKNPGVVFTAPASGK	80	90	100	110	120	
5	orf22a			нинин	111111111	111111111	1 1	
		70	80	90	100	110	120	
10	orf22.pep	130 NLIQSGLWTALRTRP						
	orf22a					VIKEAXXDFI 170	RRXXLV 180	
	The complete strain					· <del>•</del>		lan:
15		10	20	30	40	, 50	60	P
13	orf22a.pep	MIKIKKGLNLPIA	GRPEQVIYDO	SPVITEVALLG	EEYAGMRPXM	KVKEGDAVKI	KGOVLFED	
	orf22-1	MIKIKKGLNLPIA	GRPEQAVYDO 20	SPAITEVALLG 30	EEYAGMRPSM 40	KVKEGDAVKI 50	KGQVLFED 60	
20		70	80	90	100	110	120	
	orf22a.pep	KKXPGVVFTAPVS	1111111111		11111111111	111111111	111 1	
25	orf22-1	KKNPGVVFTAPAS 70	GKIAAIHRGE 80	EKRVLQSVVIA 90	VEGNDEIEFE 100	RYAPEALANI 110	LSGEEVRR 120	
	orf22a.pep	130 NLIQSGLWTALRX	140 RPFSKTPAVI	150	160 MDTND1 AADD	170	1,80	
30	orf22-1	: NLIQSGLWTALRT	1111111111	11111111111	11111111111	:1:1111	11:1 11	ŀ
	<b>V</b>	130	140	150	160	170	180	
	orf22a.pep	190 LSRLTERKIHVCK	200 AAGADVPSEN	210 NAANIETHEFG	220 GPHPAGLSGT	230 HIHFIEPVG	240 ANKTVWTI	1
35	orf22-1		AAGADVPSEN	<b>NAANIETHEFG</b>	GPHPAGLSGT	HIHFIEPVG	ANKTVWTI	
		190	200	210	220	230	240	
40	orf22a.pep	250 NYQDVIAIGRLFA       :	260 TGRLNTERVI	270 ALGGSQVNKP	280 RLLRTVLGAK	290 VSQITAGELV	300 /DADNRVI	
	orf22-1	NYQDVITIGRLFA 250	TGRLNTERVI 260	IALGGSQVNKP 270	RLLRTVLGAK 280	VSQITAGELV VSQITAGELV	/DTDNRVI 300	
45		310	320	330	340	350	360	
	orf22a.pep	SGSVLNGAITQGA	1111111111		11111111111	DKYSITRTTI	LGHFLKNK	
50	orf22-1	SGSVLNGAITQGA 310	HDYLGRYHNÇ 320	NSVIEEGRSK 330	ELFGWVAPQP 340	DKYSITRTTI 350	LGHFLKNK 360	
50	500	370	380	390	400	410	420	
	orf22a.pep orf22-1	LFKFTTAVNGGDR	11111111111		111111111	111111111	1111111	
55	01122-1	LFKFNTAVNGGDR 370	380	390	400	410	420	
	orf22a.pep	430 LCSFVCPGKYEXG	440 PLLRKVLET	KEKEGX				
60	orf22-1			11111				
٠		430	440	<b>~~~</b>			_	

Further work identified a partial gene sequence <SEQ ID 129> from N.gonorrhoeae, which encodes the following amino acid sequence <SEQ ID 130; ORF22ng>:

65

<sup>1</sup> MIKIKKGLNL PIAGRPEQVI YDGPAITEVA LLGEEYVGMR PSMKIKEGEA 51 VKKGQVLFED KKNPGVVFTA PASGKIAAIH RGEKRVLQSV VIAVEGNDEI 101 EFERYVPEAL AKLSSEKVRR NLIQSGLWTA LRTRPFSKIP AVDAEPFAIF

5

10

15

20

25

30

```
-125-
          151
               VNAMDTNPLA ADPTVIIKEA AEDFKRGLLV LSRLTERKIH VCKAAGADVP
               SENAANIETH EFGGPHPAGL SGTHIHFIEP VGANKTVWTI NYQDVIAIGR
          201
               LFVTGRLNTE RVVALGGLQV NKPRLLRTVL GAKVSQLTAG ELVDADNRVI
          251
          301
               SGSVLNGAIA QGAHDYLGRY HN*
Further work identified complete gonococcal gene <SEQ ID 131>:
               ATGATTAAAA TCAAAAAAGG TCTAAATCTG CCCATCGCGG GCAGACCGGA
               GCAAGTCATT TATGACGGCC CGGCCATTAC CGAAGTCGCG TTGCTTGGCG
           51
               AAGAATATGT CGGCATGCGC CCCTCGATGA AAATCAAGGA AGGTGAAGCC
          101
               GTCAAAAAAG GCCAAGTGCT GTTTGAAGAC AAAAAGAATC CGGGCGTAGT
          151
               ATTTACTGCG CCGGCTTCAG GCAAAATCGC CGCTATTCAC CGTGGCGAAA
          201
               AGCGCGTACT TCAGTCAGTC GTGATTGCCG TTGAAGGCAA CGACGAAATC
          251
               GAGTTCGAAC GCTACGTACC TGAAGCGCTG GCAAAATTGA GCAGCGAAAA
          301
               AGTGCGCCGC AACCTGATTC AATCAGGCTT ATGGACTGCG CTTCGCACCC
          351
               GTCCGTTCAG CAAAATCCCT GCCGTAGATG CCGAGCCGTT CGCCATCTTC
          401
               GTCAATGCGA TGGACACCAA TCCGCTGGCT GCCGACCCTA CGGTCATCAT
          451
               CAAAGAAGCC GCCGAAGACT TCAAACGCGG CCTGTTGGTA TTGAGCCGCC
          501
               TGACCGAACG TAAAATCCAT GTGTGTAAAG CAGCAGGCGC AGACGTGCCG
          551
               TCTGAAAATG CTGCCAATAT CGAAACACAT GAATTTGGCG GCCCGCATCC
          601
          651
               TGCCGGCTTG AGTGGCACGC ACATTCATTT CATCGAGCCA GTCGGCGCGA
               ATAAAACCGT GTGGACCATC AATTATCAAG ACGTGATTGC TATCGGACGT
          701
               TTGTTCGTAA CAGGCCGTCT GAATACCGAG CGCGTGGTTG CCTTGGGCGG
          751
          801
               CCTGCAAGTC AACAAACCGC GCCTCTTGCG TACCGTTTTG GGTGCGAAGG
               TGTCTCAACT TACCGCCGGC GAATTGGTTG ACGCGGACAA CCGCGTGATT
          851
               TCCGGTTCGG TATTGAACGG TGCGATTGCA CAAGGCGCGC ATGATTATTT
          901
               GGGACGCTAC CACAATCAGA TTTCCGTTAT CGAAGAAGGC CGCAGCAAAG
          951
         1001
               AGCTGTTCGG CTGGGTTGCG CCGCAGCCGG ACAAATACTC CATCACGCGC
               ACCACTCTCG GCCATTTCCT AAAAAACAAA CTCTTCAAGT TCACGACAGC
         1051
         1101
               CGTCAACGGC GGCGACCGCG CCATGGTACC GATCGGCACT TATGAGCGCG
               TAATGCCGTT GGACATCCTG CCTACCTTGC TTTTGCGCGA TTTAATCGTC
```

This encodes a protein having amino acid sequence <SEQ ID 132; ORF22ng-1>:

CGCTGTTGCG CAAAGTGCTG GAAACCATTG AGAAGGAAGG CTGA

```
MIKIKKGLNL PIAGRPEQVI YDGPAITEVA LLGEEYVGMR PSMKIKEGEA
35
                    VKKGQVLFED KKNPGVVFTA PASGKIAAIH RGEKRVLQSV VIAVEGNDEI
                51
               101
                    EFERYVPEAL AKLSSEKVRR NLIQSGLWTA LRTRPFSKIP AVDAEPFAIF
               151
                    VNAMDTNPLA ADPTVIIKEA AEDFKRGLLV LSRLTERKIH VCKAAGADVP
                    SENAANIETH EFGGPHPAGL SGTHIHFIEP VGANKTVWTI NYQDVIAIGR
               201
                    LFVTGRLNTE RVVALGGLQV NKPRLLRTVL GAKVSQLTAG ELVDADNRVI
               251
40
               301
                    SGSVLNGAIA QGAHDYLGRY HNQISVIEEG RSKELFGWVA PQPDKYSITR
                    TTLGHFLKNK LFKFTTAVNG GDRAMVPIGT YERVMPLDIL PTLLLRDLIV
               351
                    GDTDSAQALG CLELDEEDLA LCSFVCPGKY EYGPLLRKVL ETIEKEG*
               401
```

The originally-identified partial strain B sequence (ORF22) shows 93.7% identity over a 158aa

GGCGATACCG ACAGCGCGCA GGCTTTGGGT TGCTTGGAAT TGGACGAAGA AGACCTCGCT TTGTGCAGCT TCGTCTGCCC GGGCAAATAC GAATACGGCC

overlap with ORF22ng: 45

1151

1201

1251

	orf22.pep	MIKIKKGLNLPIAGRPEQAVYDGPAITEVALLGEEYAGMRPSMKVKEGDAVKKGQVLFED	60
•	orf22ng		60
50	orf22.pep	KKNPGVVFTAPASGKIAAIHRGEKRVLQSVVIAVEXNDEIEFERYAPEALANLSGEEVRR	120
	orf22ng	KKNPGVVFTAPASGKIAAIHRGEKRVLQSVVIAVEGNDEIEFERYVPEALAKLSSEKVRR	. 120
55	orf22.pep	NLIQSGLWTALRTRPFSKIPAVDAEPFAIFVNAMDTNP	158
	orf22ng	NLIQSGLWTALRTRPFSKIPAVDAEPFAIFVNAMDTNPLAADPTVIIKEAAEDFKRGLLV	180

The complete sequences from strain B (ORF22-1) and gonococcus (ORF22ng) show 96.2% identity in 447 aa overlap:

60 50 60 orf22-1.pep MIKIKKGLNLPIAGRPEQAVYDGPAITEVALLGEEYAGMRPSMKVKEGDAVKKGQVLFED -126-

		1111111	11111111111	1::111111	1111111111	:	11:111111	1111
	orf22ng-1					VGMRPSMKIK		
	Ollering 1		10	20	30	40	50	60
			10	20	30	40	30	00
•			20	80	00	100		
5			.70		90	100	110	120
	orf22-1.pep					NDE IEFERYA:		
						111111111:		
	orf22ng-1	KKNPGVV	ftapasgki <i>f</i>	AIHRGEKRV	LQSVVIAVEG	NDEIEFERYV:	PEALAKLSSE	KVRR
			70	80	90	100	110	120
10								
••			130	140	150	160	170	180
	orf22-1.pep					NPLAADPTVI:		
	Office 1. pep	NEIQUOE	111111111	LILILIII	IIIIIIIIIII	111111111	INCAREDERA	CTTA
	500 1	111111				1111111111	111111111	1111
1.5	orf22ng-1					NPLAADPTVI		
15			130	140	150	160	170	180
			190	200	210	220	230	240
	orf22-1.pep	LSRLTER	KIHVCKAAGA	DVPSENAAN	IETHEFGGPH	PAGLSGTHIH:	FIEPVGANKT	'VWTI
		1111111	1111111111	111111111	111111111		шийш	1111
20	orf22ng-1	LSRITER	KTHVCKAAGA	DVPSENAAN	TETHERGORE	PAGLSGTHIH:	FTFDVCANKT	ነ ነ ነ ነ ፕ <b>ም</b> ሞፕ
20	Office in a second		190	200	210	220	230	240
			170	200	210	220	230	240
			250	200	220	000	000	
			250	260	270	280	290	300
	orf22-1.pep					RTVLGAKVSQ		
25						1111111111		
	orf22ng-1	NYQDVIA	.IGRLFVTGRI	NTERVVALG	GLQVNKPRLL	RTVLGAKVSQ:	LTAGELVDAD	NRVI
			250	260	270	280	290	300
							,	
			310	320	330	340	350	360
30	orf22-1.pep					GWVAPQPDKY:		
50	Office 1.pep							
						GWVAPQPDKY:		
	orf22ng-1	2G2 A PING				_		
			310	320	330	. 340	350	360
0.5								
35			370	380	390	400	410	420
	orf22-1.pep	LFKFNTA	VNGGDRAMVI	PIGTYERVMP	LDILPTLLLR	DLIVGDTDSA	QALGCLELDE	EDLA
		1111:11	1111111111		1111111111	1111111111	1111111111	1111
	orf22ng-1	LFKFTTA	VNGGDRAMV	IGTYERVMP	LDILPTLLLR	DLIVGDTDSA	OALGCLELDE	EDLA
	- "		370	380	390	400	410	420
40			- · •	<del>-</del>				
70			430	440				
	omf22_1 non	T COESICE	GKYEYGPLLE		cv			
	orf22-1.pep							
4.5	orf22ng-1	LCSFVCE	GKYEYGPLL		GX			
45			430	440				

Computer analysis of these sequences gave the following results:

Homology with 48kDa outer membrane protein of *Actinobacillus pleuropneumoniae* (accession number U24492). ORF22 and this 48kDa protein show 72% aa identity in 158aa overlap:

50	Orf22	1	MIKIKKGLNLPIAGRPEQAVYDGPAITEVALLGEEYAGMRPSMKVKEGDAVKKGQVLFED 60 MI IKKGL+LPIAG P Q +++G + EVA+LGEEY GMRPSMKV+EGD VKKGOVLFED
	48kDa	1	MITIKKGLDLPIAGTPAQVIHNGNTVNEVAMLGEEYVGMRPSMKVREGDVVKKGQVLFED 60
	orf22	61	KKNPGVVFTAPASGKIAAIHRGEKRVLQSVVIAVEXNDEIEFERYAPEALANLSGEEVRR 120
55	48kDa	61	KKNPGVVFTAPASG + I+RGEKRVLQSVVI VE +++I F RY LA+LS E+V++ KKNPGVVFTAPASGTVVTINRGEKRVLQSVVIKVEGDEQITFTRYEAAQLASLSAEQVKQ 120
	orf22	121	NLIQSGLWTALRTRPFSKIPAVDAEPFAIFVNAMDTNP 158 NLI+SGLWTA RTRPFSK+PA+DA P +IFVNAMDTNP
60	48kDa	121	NLIESGLWTAFRTRPFSKVPALDAIPSSIFVNAMDTNP 158

ORF22a also shows homology to the 48kDa Actinobacillus pleuropneumoniae protein:

```
gi|1185395 (U24492) 48 kDa outer membrane protein [Actinobacillus pleuropneumoniae] Length = 449
```

Score = 530 bits (1351), Expect = e-150

```
Identities = 274/450 (60%), Positives = 323/450 (70%), Gaps = 4/450 (0%)
          Query: 1
                     MIKIKKGLNLPIAGRPEQVIYDGPVITEVALLGEEYAGMRPXMKVKEGDAVKKGQVLFED 60
                     MI IKKGL+LPIAG P QVI++G + EVA+LGEEY GMRP MKV+EGD VKKGOVLFED
5
          Sbjct: 1
                     MITIKKGLDLPIAGTPAQVIHNGNTVNEVAMLGEEYVGMRPSMKVREGDVVKKGQVLFED 60
          Query: 61
                     KKXPGVVFTAPVSGKIAAIHRGEKRVLQSVVIAVEGNDEIEFERYAPEALANLSGXEXXX 120
                     KK PGVVFTAP SG + I+RGEKRVLQSVVI VEG+++I F RY
          Sbjct: 61 KKNPGVVFTAPASGTVVTINRGEKRVLQSVVIKVEGDEQITFTRYEAAQLASLSAEQVKQ 120
10
          Query: 121 NLIQSGLWTALRXRPFSKIPAVDAEPFAIFVNAMDTNPLAADPVVVIKEAXXDFRRXXLV 180
                     NLI+SGLWTA R RPFSK+PA+DA P +IFVNAMDTNPLAADP VV+KE
                                                                      DF+
          Sbjct: 121 NLIESGLWTAFRTRPFSKVPALDAIPSSIFVNAMDTNPLAADPEVVLKEYETDFKDGLTV 180
15
          Query: 181 LSRL--TERKIHVCKAAGADVP-SENAANIETHEFGGPHPAGLSGTHIHFIEPVGANKTV 237
                     L+RL ++ +++CK A +++P S I F G HPAGL GTHIHF++PVGA K V
          Sbjct: 181 LTRLFNGQKPVYLCKDADSNIPLSPAIEGITIKSFSGVHPAGLVGTHIHFVDPVGATKQV 240
          Query: 238 WTINYQDVIAIGRLFATGRLNTERVIALGGSQVNKPRLLRTVLGAKVSQITAGELVDADN 297
                     W +NYQDVIAIG+LF TG L T+R+I+L G QV PRL+RT LGA +SQ+TA EL
20
          Sbjct: 241 WHLNYQDVIAIGKLFTTGELFTDRIISLAGPQVKNPRLVRTRLGANLSQLTANELNAGEN 300
          Query: 298 RVISGSVLNGAITQGAHDYLGRYHNQISVIEEGRSKELFGWVAPQPDKYSITRTTLGHFL 357
                     RVISGSVL+GA G DYLGRY Q+SV+ EGR KELFGW+ P DK+SITRT LGHF
25
          Sbjct: 301 RVISGSVLSGATAAGPVDYLGRYALQVSVLAEGREKELFGWIMPGSDKFSITRTVLGHFG 360
          K KLF FTTAV+GG+RAMVPIG YERVM
                                                               GDTDSAO
          Sbjct: 361 K-KLFNFTTAVHGGERAMVPIGAYERVMPLDIIPTLLLRDLAAGDTDSAONLGCLELDEE 419
30
          Query: 418 XXXXXSFVCPGKYEXGPLLRKVLETXEKEG 447
                          ++VCPGK
                                   GP+LR LE EKEG
     ORF22ng-1 also shows homology with the OMP from A.pleuropneumoniae:
                        (U24492)
                                   48
                                       kDa
                                             outer
                                                     membrane
                                                                protein
                                                                          [Actinobacillus
35
          pleuropneumoniae | Length = 449
           Score = 555 bits (1414), Expect = e-157
           Identities = 284/450 (63%), Positives = 337/450 (74%), Gaps = 4/450 (0%)
          Query: 27 MIKIKKGLNLPIAGRPEQVIYDGPAITEVALLGEEYVGMRPSMKIKEGEAVKKGQVLFED 86
40
                     MI IKKGL+LPIAG P QVI++G + EVA+LGEEYVGMRPSMK++EG+ VKKGOVLFED
                     MITIKKGLDLPIAGTPAQVIHNGNTVNEVAMLGEEYVGMRPSMKVREGDVVKKGQVLFED 60
          Sbjct: 1
                     KKNPGVVFTAPÄSGKIAAIHRGEKRVLQSVVIAVEGNDEIEFERYVPEALAKLSSEKVRR 146
          Query: 87
                     KKNPGVVFTAPASG + I+RGEKRVLQSVVI VEG+++I F RY
45
          Sbjct: 61 KKNPGVVFTAPASGTVVTINRGEKRVLQSVVIKVEGDEQITFTRYEAAQLASLSAEQVKQ 120
          Query: 147 NLIQSGLWTALRTRPFSKIPAVDAEPFAIFVNAMDTNPLAADPTVIIKEAAEDFKRGLLV 206
                     NLI+SGLWTA RTRPFSK+PA+DA P +IFVNAMDTNPLAADP V++KE
          Sbjct: 121 NLIESGLWTAFRTRPFSKVPALDAIPSSIFVNAMDTNPLAADPEVVLKEYETDFKDGLTV 180
50
          Query: 207 LSRL--TERKIHVCKAAGADVP-SENAANIETHEFGGPHPAGLSGTHIHFIEPVGANKTV 263
                     L+RL ++ +++CK A +++P S
                                                I
                                                      F G HPAGL GTHIHF++PVGA K V
          Sbjct: 181 LTRLFNGQKPVYLCKDADSNIPLSPAIEGITIKSFSGVHPAGLVGTHIHFVDPVGATKQV 240
55
          Query: 264 WTINYQDVIAIGRLFVTGRLNTERVVALGGLQVNKPRLLRTVLGAKVSQLTAGELVDADN 323
                     W +NYQDVIAIG+LF TG L T+R+++L G QV PRL+RT LGA +SQLTA EL
           Sbjct: 241 WHLNYQDVIAIGKLFTTGELFTDRIISLAGPQVKNPRLVRTRLGANLSQLTANELNAGEN 300
           Query: 324 RVISGSVLNGAIAQGAHDYLGRYHNQISVIEEGRSKELFGWVAPQPDKYSITRTTLGHFL 383
60
                     RVISGSVL+GA A G DYLGRY Q+SV+ EGR KELFGW+ P DK+SITRT LGHF
           Sbjct: 301 RVISGSVLSGATAAGPVDYLGRYALQVSVLAEGREKELFGWIMPGSDKFSITRTVLGHFG 360
           Query: 384 KNKLFKFTTAVNGGDRAMVPIGTYERVMXXXXXXXXXXXXXXVGDTDSAQXXXXXXXXXX 443
                     K KLF FTTAV+GG+RAMVPIG YERVM
                                                               GDTDSAO
65
           Sbjct: 361 K-KLFNFTTAVHGGERAMVPIGAYERVMPLDIIPTLLLRDLAAGDTDSAONLGCLELDEE 419
           Query: 444 XXXXXSFVCPGKYEYGPLLRKVLETIEKEG 473
                          ++VCPGK YGP+LR LE IEKEG
           Sbjct: 420 DLALCTYVCPGKNNYGPMLRAALEKIEKEG 449
70
```

Based on this analysis, including the homology with the outer membrane protein of *Actinobacillus* pleuropneumoniae, it was predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF22-1 (35.4kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 5A shows the results of affinity purification of the GST-fusion protein, and Figure 5B shows the results of expression of the His-fusion in *E.coli*. Purified GST-fusion protein was used to immunise mice, whose sera were used for ELISA (positive result) and FACS analysis (Figure 5C). These experiments confirm that ORF22-1 is a surface-exposed protein, and that it is a useful immunogen.

#### 10 Example 16

5

The following partial DNA sequence was identified in N.meningitidis <SEQ ID 133>:

```
..GCGnCGnAAA TCATCCATCC CC..nACGTC GTAGGCCCTG AAGCCAACTG
                 51
                       GTTTTTTATG GTAGCCAGTA CGTTTGTGAT TGCTTTGATT GGTTATTTTG
                       TTACTGAAAA AATCGTCGAA CCGCAATTGG GCCCTTATCA ATCAGATTTG
                101
15
                151
                       TCACAAGAAG AAAAAGACAT TCGGCATTCC AATGAAATCA CGCCTTTGGA
                       ATATAAAGGA TTAATTTGGG CTGGCGTGGT GTTTGTTGCC TTATCCGCCC
                201
                251
                       TATTGGCTTG GAGCATCGTC CCTGCCGACG GTATTTTGCG TCATCCTGAA
                       ACAGGATTGG TTTCCGGTTC GCCGTTTTTA AAATCGATTG TTGTTTTTAT
                301
                351
                       TTTCTTGTTG TTTGCACTGC CGGGCATTGT TTATGGCCGG GTAACCCGAA
20
                401
                       GTTTGCGCGG CGAACAGGAA GTCGTTAATG CGmyGGCCGA ATCGATGAGT
                451
                       ACTCTGGsGC TTTmTTTGsw CAkcATCTTT TTTGCCGCAC AGTTTGTCGC
                501
                       ATTTTTTAAT TGGACGAATA TTGGGCAATA TATTGCCGTT AAAGGGGCGA
                551
                       CGTTCTTAAA AGAAGTCGGC TTGGGCGGCA GCGTGTTGTT TATCGGTTTT
                601
                       ATTTTAATTT GTGCTTTTAT CAATCTGATG ATAGGCTCCG CCTCCGCGCA
25
                651
                       ATGGGCGGTA ACTGCGCCGA TTTTCGTCCC TATGCTGATG TTGGCCGGCT
                       ACGCGCCCGA AGTCATTCAA GCCGCTTACC GCATCGGTGA TTCCGTTACC
                701
                751
                       AATATTATTA CGCCGATGAT GAGTTATTTC GGGCTGATTA TGGCGACGGT
                       GrkCmmmTAC AAAAAAGATG CGGGCGTGGG TaCGcTGATT wCTATGATGT
                801
                851
                       TGCCGTATTC CGCTTTCTTC TTGATTGCgT GGATTGCCTT ATTCTGCATT
30
                       TGGGTATTTG TTTTGGGCCT GCCCGTCGGT CCCGGCGCGC CCACATTCTA
                901
                       TCCCGCACCT TAA
```

This corresponds to the amino acid sequence <SEQ ID 134; ORF12>:

```
1 ..AXXIIHPXXV VGPEANWFFM VASTFVIALI GYFVTEKIVE PQLGPYQSDL
51 SQEEKDIRHS NEITPLEYKG LIWAGVVFVA LSALLAWSIV PADGILRHPE
35 101 TGLVSGSPFL KSIVVFIFLL FALPGIVYGR VTRSLRGEQE VVNAXAESMS
151 TLXLXXXIF FAAQFVAFFN WTNIGQYIAV KGATFLKEVG LGGSVLFIGF
201 ILICAFINLM IGSASAQWAV TAPIFVPMLM LAGYAPEVIQ AAYRIGDSVT
251 NIITPMMSYF GLIMATVXXY KKDAGVGTLI XMMLPYSAFF LIAWIALFCI
301 WVFVLGLPVG PGAPTFYPAP *
```

40 Further sequence analysis revealed the complete DNA sequence <SEQ ID 135> to be:

```
ATGAGTCAAA CCGATACGCA ACGGGACGGA CGATTTTTAC GCACAGTCGA
                 1
                51
                    ATGGCTGGGC AATATGTTGC CGCATCCGGT TACGCTTTTT ATTATTTTCA
                    TTGTGTTATT GCTGATTGCC TCTGCCGTCG GTGCGTATTT CGGACTATCC
               101
               151
                    GTCCCCGATC CGCGCCCTGT TGGTGCGAAA GGACGTGCCG ATGACGGTTT
45
                    GATTTACATT GTCAGCCTGC TCAATGCCGA CGGTTTTATC AAAATCCTGA
               201
               251
                    CGCATACCGT TAAAAATTTC ACCGGTTTCG CGCCGTTGGG AACGGTGTTG
               301
                    GTTTCTTTAT TGGGCGTGGG GATTGCGGAA AAATCGGGCT TGATTTCCGC
                    ATTAATGCGC TTATTGCTCA CAAAATCGCC ACGCAAACTC ACTACTTTTA
               351
                    TGGTTGTTTT TACAGGGATT TTATCTAATA CCGCTTCTGA ATTGGGCTAT
               401
50
               451
                    GTCGTCCTAA TCCCTTTGTC CGCCATCATC TTTCATTCCC TCGGCCGCCA
                    TCCGCTTGCC GGTCTGGCTG CGGCTTTCGC CGGCGTTTCG GGCGGTTATT
               501
```

	551	CGGCCAATCT	GTTCTTAGGC	ACAATCGATC	CGCTCTTGGC	AGGCATCACC
	601	CAACAGGCGG	CGCAAATCAT	CCATCCCGAC	TACGTCGTAG	GCCCTGAAGC
	651	CAACTGGTTT	TTTATGGTAG	CCAGTACGTT	TGTGATTGCT	TTGATTGGTT
	701	ATTTTGTTAC	TGAAAAAATC	GTCGAACCGC	AATTGGGCCC	TTATCAATCA
5	751	GATTTGTCAC	AAGAAGAAAA	AGACATTCGG	CATTCCAATG	AAATCACGCC
	801	TTTGGAATAT	AAAGGATTAA	TTTGGGCTGG	CGTGGTGTTT	GTTGCCTTAT
	851	CCGCCCTATT	GGCTTGGAGC	ATCGTCCCTG	CCGACGGTAT	TTTGCGTCAT
•	901	CCTGAAACAG	GATTGGTTTC	CGGTTCGCCG	TTTTTAAAAT	CGATTGTTGT
	951	TTTTATTTTC	TTGTTGTTTG	CACTGCCGGG	CATTGTTTAT	GGCCGGGTAA
10	1001	CCCGAAGTTT	GCGCGGCGAA	CAGGAAGTCG	TTAATGCGAT	GGCCGAATCG
•	1051	ATGAGTACTC	TGGGGCTTTA	TTTGGTCATC	ATCTTTTTTG	CCGCACAGTT
	1101	TGTCGCATTT	TTTAATTGGA	CGAATATTGG	GCAATATATT	GCCGTTAAAG
	1151	GGGCGACGTT	CTTAAAAGAA	GTCGGCTTGG	GCGGCAGCGT	GTTGTTTATC
	1201	GGTTTTATTT	TAATTTGTGC	TTTTATCAAT	CTGATGATAG	GCTCCGCCTC
15	1251	CGCGCAATGG	GCGGTAACTG	CGCCGATTTT	CGTCCCTATG	CTGATGTTGG
	1301	CCGGCTACGC	GCCCGAAGTC	ATTCAAGCCG	CTTACCGCAT	CGGTGATTCC
	1351	GTTACCAATA	TTATTACGCC	GATGATGAGT	TATTTCGGGC	TGATTATGGC
	1401			AAGATGCGGG	CGTGGGTACG	CTGATTTCTA
	1451		GTATTCCGCT		TTGCGTGGAT	
20	1501			GGGCCTGCCC	GTCGGTCCCG	GCGCGCCCAC
	1551	ATTCTATCCC	GCACCTTAA			•

This corresponds to the amino acid sequence <SEQ ID 136; ORF12-1>:

	. 1	MSQTDTQRDG	RFLRTVEWLG	NMLPHPVTLF	IIFIVLLLIA	SAVGAYFGLS
	51	VPDPRPVGAK	GRADDGLIYI	VSLLNADGFI	KILTHTVKNF	TGFAPLGTVL
25	101	VSLLGVGIAE	KSGLISALMR	LLLTKSPRKL	TTFMVVFTGI	LSNTASELGY
	151	VVLIPLSAII	FHSLGRHPLA	GLAAAFAGVS	GGYSANLFLG	TIDPLLAGIT
	201	QQAAQIIHPD	YVVGPEANWF	FMVASTFVIA	LIGYFVTEKI	VEPQLGPYQS
	251	DLSQEEKDIR	HSNEITPLEY	KGLIWAGVVF	VALSALLAWS	IVPADGILRH
	301	PETGLVSGSP	FLKSIVVFIF	LLFALPGIVY	GRVTRSLRGE	QEVVNAMAES
30	351	MST <u>LGLYLVI</u>	IFFAAQFVAF	FNWTNIGQYI	AVKGATFLKE	VGLGGSVLFI
	401	GFILICAFIN	LMIGSASAQW	AVTAPIFVPM	LMLAGYAPEV	IQAAYRIGDS
	451	VTNIITPMMS	YFGLIMATVI	KYKKDAGVGT	LISMMLPYSA	FFLIAWIALF
	501	CIWVFVLGLP	VGPGAPTFYP	AP*		

Computer analysis of this amino acid sequence gave the following results:

# 35 Homology with a predicted ORF from N. meningitidis (strain A)

ORF12 shows 96.3% identity over a 320aa overlap with an ORF (ORF12a) from strain A of N. meningitidis:

	*				. 10	20	30
• • •	orf12.pep			AXX	XIIHPXXVVG	Peanwffmva:	STFVIALI
40	*			1	1111 111		[[]]
	orf12a	AAAFAGVSGGY:		PLLAGITQQAA	QIIHPDYVVG	PEANWFFMVA:	STFVIALI
		180	190	200	210	220	230
			50	60	70		
45		40	50	60	70	80	90
43	orf12.pep	GYFVTEKIVEP	Tresidantad	EEKDIKHSNE.	TIPLEYKGLI	WAGVVEVALS	
	610.				111111111	111111111	11111111
	orf12a	GYFVTEKIVEP		•			
		240	250	260	270	280	290
50	•	100	110	120	130	140	150
	orf12.pep	PADGILRHPET					
	OIIII.POP	· 1111111111	131111111111		111111111	1111111111	
	orf12a	PADGILRHPET	ST.VSGSPFT.KS		I.PGTVYCRVT		
	011124	300	310	320	330	340	350
55		300	310	320	330	340	330
55		160	170	180	190	200	210
	orf12.pep	TLXLXLXXIFF			-		
	OLLIE, POP				111111111	1111111111	IIIIIIIIII
•	orf12a	TLGLYLVIIFF		NTGOYTAVKG	ATFLKEVCLC	GSVI.FTGFTI.	
60	V	360	370	380	390	400	410
00	•	300	3.0	300	330	100	410
		220	230	240	250	260	270
	orf12.pep	IGSASAQWAVT	API FVPMLML#	GYAPEVIOAA	YRIGDSVTNI	ITPMMSYFGI.	
		HIHITI			1111111111		11111

	orf12a	IGSASAQW 420		MLAGYAPEVI( 440	QAAYRIGDSVTN 450	NIITPMMSYFGL 460	MATVIKY 470
_	510		80 29			320	
5	orf12.pep				[WVFVLGLPVGI		
	orf12a		LISMMLPYSAR		WVFVLGLPVGI 510		
	The complete les	ngth ORF12a	nucleotide :	sequence <s< th=""><th>EQ ID 137&gt;</th><th>is:</th><th></th></s<>	EQ ID 137>	is:	
10	1	<b>አ</b> ምሮ <b>እ</b> ሮምሮአ አ አ	CCCATTACCCA	ACCCCACCCA	CGATTTTTAC	CC) C) CECC)	
10	51	ATGGCTGGGC	AATATGTTGC	CGCACCCGGT	TACGCTTTTT	ATTATTTTCA	
	101	TTGTGTTATT	GCTGATTGCC	TCTGCCGCCG	GTGCGTATTT	CGGACTATCC	
	151				GGACGTGCCG		
15	201 251				CGCCGTTGGG		
13	301	GTTTCTTTAT	TGGGCGTGGG	GATTGCGGAA	AAATCGGGCT	TGATTTCCGC	
	351	ATTAATGCGC	TTATTGCTCA	CAAAATCTCC	ACGCAAACTC	ACTACTTTTA	
	401				CCGCTTCTGA		
20	451 501				TTTCATTCCC CGGCGTTTCG		
20	551	CGGCCAATCT	GTTCTTAGGC	ACAATCGATC	CGCTCTTGGC	AGGCATCACC	
	601	CAACAGGCGG	CGCAAATCAT	CCATCCCGAC	TACGTCGTAG	GCCCTGAAGC	
	651				TGTGATTGCT		
25	701 751				AATTGGGCCC CATTCCAATG	TTATCAATCA	
23	801	TTTGGAATAT	AAAGGATTAA	TTTGGGCTGG	CGTGGTGTTT	GTTGCCTTAT	;
	851	CCGCCCTATT	GGCTTGGAGC	ATCGTCCCTG	CCGACGGTAT	TTTGCGTCAT	•
	901	CCTGAAACAG	GATTGGTTTC	CGGTTCGCCG	TTTTTAAAAT	CAATTGTTGT	
30	951 1001				CATTGTTTAT TTAATGCGAT		
50	1051				ATCTTTTTTG		
	1101	TGTCGCATTT	TTTAATTGGA	<b>CGAATATTGG</b>	GCAATATATT	GCCGTTAAAG	
	1151	GGGCGACGTT	CTTAAAAGAA	GTCGGCTTGG	GCGGCAGCGT	GTTGTTTATC	
35	1201 1251	CCCCCAATCC	GCGGTAACTG	CCCCATTT	CTGATGATAG CGTCCCTATG	GCTCCGCCTC	
55	1301	CCGGCTACGC	GCCCGAAGTC	ATTCAAGCCG	CTTACCGCAT	CGGTGATTCC	
	1351	GTTACCAATA	TTATTACGCC	GATGATGAGT	TATTTCGGGC	TGATTATGGC	
	1401	GACGGTGATC	AAATACAAAA	AAGATGCGGG	CGTGGGTACG	CTGATTTCTA	
40	1451 1501	TGATGTTGCC	GTATTCCGCT TATTTCTTTT	GCCCTCCC	TTGCGTGGAT GTCGGTCCCG	TGCCTTATTC	
40	1551	ATTCTATCCC		0000010000	G100010000	GCGCGCCAC	
	This encodes a p	protein having	g amino acid	sequence <	SEQ ID 138	>:	
	1	MSQTDTQRDG	RFLRTVEWLG	NMLPHPVTLF	IIFIVLLLIA	SAAGAYFGLS	
4.5	51				KILTHTVKNF		
45	101 151	VSLLGVGIAE	KSGLISALMR	CLARAFACUE	TTFMVVFTGI GGYSANLFLG	LSNTASELGY	
	201				LIGYFVTEKI		
	251	DLSQEEKDIR	HSNEITPLEY	KGLIWAGVVF	VALSALLAWS	IVPADGILRH	
50	301				GRVTRSLRGE		
50	351 401	GEILICATIN	INTESASAOW	FUMINICOLI	AVKGATFLKE LMLAGYAPEV	VGLGGSVLFI	
	451	VTNIITPMMS	YFGLIMATVI	KYKKDAGVGT	LISMMLPYSA	FFLIAWIALF	
	501	CIWVFVLGLP	VGPGAPTFYP	AP*			
55	ORF12a and OF	RF12-1 show	99.0% ident	itv in 522 aa	overlan:		
				•			
			10	20		40 50	60
	orf12a.pe					Iasaagayfgls     :	
	orf12-1	MSOTD	ORDGRFLRTVI	EWLGNMLPHPV'	······································	:          IASAVGAYFGLS	VPDPRPVGAK
60			10	20		40 50	60
			70	80		00 110	120
	orf12a.pe	p GRADDO	ELIHVVSLLDA:	DGLIKILTHTV	KNFTGFAPLGT	VLVSLLGVGIAEI	KSGLISALMR
65	orf12-1	GRADDO	LIYIVSLLNA	OGFIKILTHTV			IIIIIIIIII KSGLISALMR
		<del>-</del> ·					

-131-

			. 70	80	90	100	110	120
	<b>63.0</b>		130	140	150	160		180
5	orf12a.pep						HPLAGLAAAFA	
	orf12-1	LLLTKSP	RKLTTFMVV	FTGILSNTAS	<b>ELGYVVLIPL</b> :	SAIIFHSLGR	HPLAGLAAAFA	.GVS
			130	140	150	160	170	180
	•		190	200	210	220		240
10	orfl2a.pep	GGYSANL	FLGTIDPLL	AGITQQAAQI	IHPDYVVGPE	ANWFFMVAST	FVIALIGYFVT	EKI
	orf12-1	GGYSANL	FLGTIDPLL	AGITOOAAOI:				III EKT
			190	200	210	220		240
15			250	260	270	280	290	300
	orf12a.pep	VEPQLGP	YQSDLSQEE	KDIRHSNEIT	PLEYKGLIWA	GVVFVALSAL	LAWSIVPADGI	LRH
	orf12-1	IIIIIII	VOCDI COEE	HIIIIIIIII	HIIIIIIII	1111111111		111
	Offiz-i		1030L3QEE. 250	260	270	3VVEVALSAL. 280		LRH 300
20		•				7.7		
	orf12a.pep		310 GSPFLKSTV	320 VFTF1.1.FA1.P	330 GTVYGRVTRS	340 Legeoevana	350 MAESMSTLGLY	360
		1111111	11111111		11111111111		1111111111	111
25	orf12-1	PETGLVS	GSPFLKSIV 310	VFIFLLFALP 320	GIVYGRVTRS: 330		MAESMSTLGLY	
23		,	310	320	330	340	350	360
	53.0		37.0	380	390	400	410	420
•	orf12a.pep	IFFAAQF	\	GQYIAVKGAT. !!!!!!!!!!	FLKEVGLGGS'	VLFIGFILIC	AFINLMIGSAS	AQW
30	orf12-1	IFFAAQF	VAFFNWTNI	GQYIAVKGAT	<b>FLKEVGLGGS</b> '	VLFIGFILIC.	AFINLMIGSAS	AQW
			370	380	390	400	410	420
			430	440	450	460	470	480
35	orf12a.pep	AVTAPIF	VPMLMLAGY:	APEVIQAAYR	IGDSVTNIIT	PMMSYFGLIM	ATVIKYKKDAG	VGT
33	orf12-1	AVTAPIF	VPMLMLAGY.	APEVIQAAYR	IGDSVTNIIT:	!!!!!!!!! PMMSYFGLIM	 ATVIKYKKDAG	 VGT
			430	440	450	460	470	480
			490	500	510	520		
40	orf12a.pep	LISMMLP	YSAFFLIAW	IALFCIWVFV	LGLPVGPGAP	TFYPAPX		
	orf12-1							
•	01112-1		15AFFLIAW 490	500	510	520		

# 45 Homology with a predicted ORF from N.gonorrhoeae

ORF12 shows 92.5% identity over a 320aa overlap with a predicted ORF (ORF12.ng) from N. gonorrhoeae:

	orf12.pep	AXXIIHPXXVVGPEANWFFMVASTFVIALI	30
50	orf12ng		232
	orf12.pep	GYFVTEKIVEPQLGPYQSDLSQEEKDIRHSNEITPLEYKGLIWAGVVFVALSALLAWSIV	90
55	orf12ng	GYFVTEKIVEPQLGPYQSDLSQEEKDIRHSNEITPLEYKGLIWAGVVFVALSALLAWSIV	292
33	orf12.pep	PADGILRHPETGLVSGSPFLKSIVVFIFLLFALPGIVYGRVTRSLRGEQEVVNAXAESMS	150
	orf12ng	PADGILRHPETGLVAGSPFLKSIVVFIFLLFALPGIVYGRITRSLRGEREVVNAMAESMS	352
60	orf12.pep	TLXLXLXXIFFAAQFVAFFNWTNIGQYIAVKGATFLKEVGLGGSVLFIGFILICAFINLM	210
	orf12ng	TLGLYLVIIFFAAQFVAFFNWTNIGQYIAVKGAVFLKKFRLGGSVLFIGFILICAFINLM	412
65	orf12.pep	IGSASAQWAVTAPIFVPMLMLAGYAPEVIQAAYRIGDSVTNIITPMMSYFGLIMATVXXY	270
	orfl2ng	IGSASAQWAVTAPIFVPMLMLAGNAPQVIQAAYRIGDSVTNIITPMMSYFGLIMATVIKY	472

	orf12.pep					IWVFVLGLPVG		320
	orf12ng							522
	The complete le	ngth OI	RF121	ng nucleotide	e sequence <	SEQ ID 139	> is:	
5	1					CGATTTTTAC		
	51					TACGCTTTTT		
	101 151					GTGCGTATTT GGACGTGCCG		
	201	GATTCA	CGTT	GTCAGCCTGC	TCGATGCCGA	CGGTTTGATC	<b>AAAATCCTGA</b>	
10	251					CGCCGTTGGG		
	301 351					AAATCGGGCT ACGCAAACTC		
	401					CGGCTTCTGA		
	451					TTTCATTCGC		
15	501 551					CGGCGTTTCG CGCTCTTGGC		
	601					TACGTCGTAG		
	651					TGTGATTGCT		
20	701					AATTGGGCCC		
20	751 801					CATTCCAATG CGTGGTGTTT		
	851					CCGACGGTAT		
	901	CCTGAA	ACAG	GATTGGTTGC	CGGTTCGCCG	TTTTTAAAAT	CGATTGTTGT	
05	951					CATTGTTTAT		
25	1001 1051					TTAATGCGAT ATCTTTTTTG		1
	1101					GCAATATATT		
	1151	GGGCGG	TGTT	CTTAAAAGAA	GTCGGCTTGG	GCGGCAGTGT	GTTGTTTATC	
30	1201					CTGATGATAG		
30	1251 1301					CGTCCCTATG CTTACCGCAT		
	1351					TATTTCGGGC		
	1401	GACGGI	PAATC	AAATACAAAA	AAGATGCGGG	CGTAGGCACG	CTGATTTCTA	
35	1451					TTGCATGGAT		
33	1501 1551			GTGCCTTAA	GGGTCTGCCC	GTCGGTCCCG	GCACACCCAC	
	This encodes a p				l seauence <	SEO ID 140	>:	
					_	-		
	1 51					IIFIVLLLIA KILTHTVKNF		
40	101					TTFMVVFTGI		
	151	VVLIPI	SAVI	FHSLGRHPLA	GLAAAFAGVS	GGYSANLFLG	TIDPLLAGIT	
	201					LIGYFVTEKI		
	251 301					VALSALLAWS GRITRSLRGE		
45	351					AVKGAVFLKK		
	401	GFILIC	CAFIN	LMIGSASAQW	AVTAPIFVPM	LMLAGNAPQV	IQAAYRIGDS	
	451	VTNIII	PMMS	YFGLIMATVI	KYKKDAGVGT	LISMMLPYSA	FFLIAWIALF	
	501	CIMAL	LGLP	VGPGTPTFYP	VP*			
	ORF12ng shows	s 97.1%	iden	tity in 522 a	a overlap wi	th ORF12-1:		
50				10	20	30	40 5	0 60
50	orf12-1.p	ep 1	4SOTD			TLFIIFIVLLL		
	•	_		::1:1111111		11111111111	11111111111	11111111111
	orf12ng	Ŋ	4SQTD					SVPDPRPVGAK
55				10	20	30	40 5	0 60
-				70	80		00 11	
	orf12-1.p					KNFTGFAPLGT		
	orf12ng					 KNFTGFAPLGT		
60	OFFIZING	,	SKADDI	70	80		VLVSLLGVGIA 00 11	
				130			60 17	
	orf12-1.p					LGYVVLIPLSA 		AGLAAAFAGVS
65	orf12ng	1	LLLTK	SPRKLTTFMVV	FTGILSNTASE	LGYVVLIPLSA	VIFHSLGRHPL	AGLAAAFAGVS
	•							

-133-

		130	140	150	160	170	180
	610 1	190	200	210	220	230	240
5	orf12-1.pep	GGYSANLFLGTIDPL	11111111	ШИНИН	111111111111111111111111111111111111111	HILLIAM	1141111
	orf12ng	GGYSANLFLGTIDPL 190	LAGITQQA. 200	AQIIHPDYVVG 210	PEANWFFMAA 220	STFVIALIG 230	YFVTEKI 240
	•	,					
10	orf12-1.pep	250 VEPQLGPYQSDLSQE	260 EKDIRHSN	270 EITPLEYKGLI	280 WAGVVFVALS	290 ALLAWSIVP	300 ADGILRH
	orf12ng	VEPQLGPYQSDLSQE	EKDTRHSN	HIIIIIIIIIII	HIIIIIIIII		IIIIIII
	VIII	250	260	270	280	290	300
15		310	320	330	340	350	360
	orf12-1.pep	PETGLVSGSPFLKSI	VVFIFLLF.	ALPGIVYGRVT	RSLRGEQEVV	NAMAESMST	LGLYLVI
	orf12ng	PETGLVAGSPFLKSI	VVFIFLLF.	ALPGIVYGRIT	RSLRGEREVV	NAMAESMST:	LGLYLVI
20		310	320	330	340	350	360
	orf12-1.pep	370 IFFAAQFVAFFNWTN	380	390	400	410	420
	• •		111111111	11:11111111	11111111111	111111111	HHIĨL
25	orf12ng	IFFAAQFVAFFNWTN 370	IIGQYIAVK 380	GAVFLKEVGLG 390	GSVLFIGFIL 400	ICAFINLMI 410	GSASAQW 420
		430	440	450	460	470	480
	orf12-1.pep	AVTAPIFVPMLMLAG	SYAPEVIQA	AYRIGDSVTNI	ITPMMSYFGL	IMATVIKYK	KDAGVGT
30	orf12ng		:          SYAPEVIQA	IIIIIIIIII AYRIGDSVTNI		  IMATVIKYK	 KDAGVGT
		430	440	450	460	470	480
	612 1	490	500	510	520		
35	orf12-1.pep	LISMMLPYSAFFLIF					
	orf12ng	LISMMLPYSAFFLIA	WIALFCIW 500	VFVLGLPVGPG 510	TPTFYPVPX 520	•	
	In addition, ORF12ng	shows significan	t homolo			rotein from	n Faoli:
40	sp P46133 YDAH >ai 1787597 (A	_ECOLI HYPOTHETI E000231) hypoth	CAL 55.1 etical p	KD PROTEIN	IN OGT-DE	PA INTERG	ENIC REGION
	Length = $510$	bits (835), Expe				(255	220 0022,
		178/507 (35%), Expe			(55%), Gap	s = 15/50	7 (2%)
45	Query: 8 RSG	RFLRTVEWLGNMLPHI	YXXXXXX	XXXXXASAVGA	YFGLSVPDPR	RPVGAKGRAD	DGL 67
	+SG	+ VE +GN +PHI KLYGWVERIGNKVPHI	?	+A+ +	FG+S +P	D	
50	Query: 68 IHV + V	VSLLDADGLIKILTH1 +LL +GL L+	VKNFTGFA KNF+GFA+	.PXXXXXXXXX .P		SALMRLLLT ALM + +	KSP 127
	Sbjct: 65 VVV	KNLLSVEGLHWFLPN	/IKNFSGFA	PLGAILALVLG	AGLAERVGLI	PALMVKMAS	HVN 124
		TTFMVVFTGILSNTAS	ELGYVVLI	PLSAVIFHSLO	RHPLAGLAAA	FAGVSGGYS.	ANL 187
55		++MV+F S+ +S ASYMVLFIAFFSHISS	S+ V++ SDAALVIMP	P+ A+IF ++G PMGALIFLAVG	RHP+AGL AA RHPVAGLLAA	A AGV G++. AIAGVGCGFT.	ANL ANL 184
		TIDPLLAGITQQAAQ1					
	+	T D LL+GI+ +AA	+P V	NW+FMA+S	V+ ++G	+T+KI+EP+	LG
60		TTDVLLSGISTEAAA					-
	Query: 248 YQS	DLSQEEKDIRHSNEI1 + ++ + + S	PLEYKGLI GI.	WAGVVFVALSA	LLAWSIVPAD		LVA 307 V
	Sbjct: 245 WQG	NSDEKLQTLTESQRF-	GLR	IAGVVSLLFIA	AIALMVIPON	GILRDPINH	TVM 298
65	Query: 308 GSP	FLKSIVVFIFLLFALI	PGIVYGRIT	RSLRGEREVVN	IAMAESMSTLO	LYLXXXXXX	XXX 367
	SP	F+K IV I L F + FIKGIVPLIILFFFV\	+ YG T	'R++R + ++ +	· M E M +	++	
						•	_
70	Query: 308 XXX	XNWTNIGQYIAVKGA NW+N+G++IAV	L+ GL	GSVLFIGFILI G F+G L+	.CAFINLMIGS - +F+ + I S	SASAQWAVTA S SA W++ A	PIF 427 PIF
				_		•	

Sbjct: 359 VAMFNWSNMGKFIAVGLTDILESSGLSGIPAFVGLALLSSFLCMFIASGSAIWSILAPIF 418 Query: 428 VPMLMLAGYAPEVIQAAYRIGDSVTNIITPMMSYFGLIMATVIKYKKDAGVGTLISMMLP 487 VPM ML G+ P Q +RI DS + P+ + L + + + YK DA +GT S++LP 5 Sbjct: 419 VPMFMLLGFHPAFAQILFRIADSSVLPLAPVSPFVPLFLGFLQRYKPDAKLGTYYSLVLP 478 Query: 488 YSAFFLIAWIALFCIWVFVLGLPVGPG 514 FL+ W+ + W +++GLP+GPG Sbict: 479 YPLIFLVVWLLMLLAW-YLVGLPIGPG 504

10

Based on this analysis, including the presence of several putative transmembrane domains and the predicted actinin-type actin-binding domain signature (shown in bold) in the gonococcal protein, it is predicted that the proteins from N. meningitidis and N. gonorrhoeae, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### 15 Example 17

The following partial DNA sequence was identified in N.meningitidis <SEO ID 141>:

```
..ACAGCCGGCG CAGCAGGTTN CNCGGTCTTC GTTTTCGTAA CGGACAGTCA
                 51
                       GGTGGAGGTG TTCGGGAACA TCCAGACCGC AGTGGAAACA GGTTTTTTTC
                101
                       ATGGCATTTC GGTTTCGTCT GTGTTTGGTG CGGCGGCACA AGACTCGGCA
20
                151
                       ATGGCTTCGC GCAGTGCGTC TATACCGGTA TTTTCAGCAA CGGAAATGCG
                201
                       GACGGCGGCA ATTTTCCCG CAGCGTCGCG CCATATGCCC GTGTTTTGTT
                251
                       CTTCAGACGG CAGCAGGTCG GTTTTGTTGT ACACCTTGAT GCACGGAaTA
                301
                       TCGCCGGCAT GGATTTCTTG CAGTACGTTT TCCACGTCTT CAATCTGCTG
                351
                       TCCGCTGTTC GGAGCGGCGG CATCGACGAC GTGCAGCAGC ACATCGGCTT
25
                401
                       gCGCGGTTTC TTCCAGCGTG GCGGAAAAGG CGGAAATCAG TTTGTGCGGC
                451
                       agATYGCTNA CGAATCCGAC GGTATCGGTC AGGATAATGC TGCATTCGGG
                501
                       ACT..
```

This corresponds to the amino acid sequence <SEQ ID 142; ORF14>:

```
.. TAGAAGXXVF VFVTDSQVEV FGNIQTAVET GFFHGISVSS VFGAAAQDSA
30
                 51
                       MASRSASIPV FSATEMRTAA IFPAASRHMP VFCSSDGSRS VLLYTLMHGI
                101
                       SPAWISCSTF STSSICCPLF GAAASTTCSS TSACAVSSSV AEKAEISLCG
                151
                       RXLTNPTVSV RIMLHSG..
```

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF14 shows 94.0% identity over a 167aa overlap with an ORF (ORF14a) from strain A of N. 35 meningitidis:

40	orf14.pep	GROLGFI	RVGGALFVT1	'AOARVNNAT	TAGAAG      :   CDCLTTGAAG		: [ : :         : [	1111
40	022230	150	160	170	180	190	200	INVE
			40	50	60	70	80	90
45	orf14.pep	GFFHGIS	VSSVFGAAAQ	DSAMASRSA	SIPVFSATEM	RTAAIFPAAS	RHMPVFCSS	DGSRS
43	orf14a	GFFHGIS	VSSVFGAAA(	IIIIIIII YSAMASRSA	SIPVFSATEM	RTAAIFPAAS	SRHMPVFCSS	DGSRS
		210	220	230	240	250	260	
			100	110	120	130	140	150
50	orfl4.pep	VLLYTLM	HGISPAWISC	STESTSSIC	CPLFGAAAST	TCSSTSACAV	/SSSVAEKAE	ISLCG
	orfl4a				CPLFGAAAST		1111111111	11111
	OTITIE	270	280	290	300	310	узээ <u>ү</u> де <u>кае</u> 320	TOPCE

268

167

orfl4ng

orfl4ng

orfl4.pep

orfl4.pep

orf14ng

RXLTNPTVSVRIMLHSG

50

55

160

```
RXLTNPTVSVRIMLHSG
          orf14.pep
                        1 111111111111111
                        RSLTNPTVSVRIMLHSGLMYSRRAVVSSVAKSWSFAYMPDLVSRLNRLDLPTLVX
          orf14a
5
                       330
                                           350
                                 340
                                                     360
                                                                          380
     The complete length ORF14a nucleotide sequence <SEQ ID 143> is:
                     ATGGAGGATT TGCAGGAAAT CGGGTTCGAT GTCGCCGCCG TAAAGGTAGG
                     TCGGCAGCGC GAACATCATC GTCTGCATCA TCCCCAGCCC GGCAACGGCG
                51
                     AGGCGGACGA TGTATTGTTT GCGTTCTTTT TGGTTGGCGG CTTCGATTTT
TTGCGCGTCA TAGGGTGCGG CGGTGTAGCC TATCTGCCTG ATTTTCAACA
                101
10
                151
                201
                     GAATGTCGGA AAGGCGGATT TTGCCGTCGT CCCAGACGAC GCGGCAGCGG
                     TGCGTGCTGT AATTGAGGTC GATGCGGACG ATGCCGTCTG TACGCAAAAG
                251
                301
                     CTGCTGTTCG ATCAGCCAGA CGCAGGCGGC GCAGGTGATG CCGCCGAGCA
                     TTAAAACCGC CTCGCGCGTG CCGCCGTGGG TTTCCACAAA GTCGGACTGG
                351
15
                401
                     ACTTCGGGCA GGTCGTACAG GCGGATTTGG TCGAGGATTT CTTGGGGCGG
                     CAGCTCGGTT TTTTGCGCGT CGGCGGTGCG TTGTTTGTAA TAACTGCCCA
                451
                501
                     AGCCCGCGTC AATAATGCTT TGTGCGACTG CCTGACAACC GGCGCAGCAG
                551
                     GTTTCGCGGT CTTCGTTTTC GTAACGGACG GTCAGATGCA GGTTTTCGGG
                     AACGTCCAGC CCGCAGTGGA AACAGGTTTT TTTCATGGCA TTTCGGTTTC
                601
20
                651
                     GTCTGTGTTT GGTGCGGCGG CACAATACTC GGCAATGGCT TCGCGCAGTG
                     CGTCTATACC GGTATTTTCA GCAACGGAAA TGCGGACGGC GGCAATTTTT
                701
                     CCCGCAGCGT CGCGCCATAT GCCCGTGTTT TGTTCTTCAG ACGGCAGCAG
                751
                     GTCGGTTTTG TTGTACACCT TGATGCACGG AATATCGCCG GCATGGATTT
                801
                851
                     CTTGCAGTAC GTTTTCCACG TCTTCAATCT GCTGTCCGCT GTTCGGAGCG
25
                     GCGGCATCGA CGACGTGCAG CAGCACATCG GCTTGCGCGG TTTCTTCCAG
                901
                     CGTGGCGGAA AAGGCGGAAA TCAGTTTGTG CGGCAGATCG CTGACGAATC
                951
               1001
                     CGACGGTATC GGTCAGGATA ATGCTGCATT CGGGACTGAT GTACAGCCGC
                     CGCGCCGTCG TGTCGAGTGT GGCGAAAAGC TGGTCTTTCG CATATATGCC
               1051
                     CGACTTGGTC AGCCGGTTGA ACAGACTGGA TTTGCCGACA TTGGTATAG
               1101
30
     This encodes a protein having amino acid sequence <SEO ID 144>:
                     MEDLQEIGFD VAAVKVGRQR EHHRLHHPQP GNGEADDVLF AFFLVGGFDF
                     LRVIGCGGVA YLPDFQQNVG KADFAVVPDD AAAVRAVIEV DADDAVCTQK
                101
                     LLFDQPDAGG AGDAAEH*NR LARAAVGFHK VGLDFGQVVQ ADLVEDFLGR
                     QLGFLRVGGA LFVITAQARV NNALCDCLTT GAAGFAVFVF VTDGQMQVFG
35
                201
                     NVQPAVETGF FHGISVSSVF GAAAQYSAMA SRSASIPVFS ATEMRTAAIF
                251
                     PAASRHMPVF CSSDGSRSVL LYTLMHGISP AWISCSTFST SSICCPLFGA
                301
                     AASTTCSSTS ACAVSSSVAE KAEISLCGRS LTNPTVSVRI MLHSGLMYSR
                351
                     RAVVSSVAKS WSFAYMPDLV SRLNRLDLPT LV*
     It should be noted that this sequence includes a stop codon at position 118.
40
     Homology with a predicted ORF from N.gonorrhoeae
      ORF14 shows 89.8% identity over a 167aa overlap with a predicted ORF (ORF14.ng) from N.
      gonorrhoeae:
           orfl4.pep
                                                      TAGAAGXXVFVFVTDSQVEVFGNIQTAVET
                                                                                        30
                                                       11 111
                                                              11:11:1:1::1111:1 1111
45
                        GRQFGFFRVGGASFVITAQAGIDDALCDCLTADAAGFAVFAFVADGQMQVFGNVQPAVET
           orfl4ng
                                                                                       208
           orfl4.pep
                        GFFHGISVSSVFGAAAQDSAMASRSASIPVFSATEMRTAAIFPAASRHMPVFCSSDGSRS
                                                                                        90
```

The complete length ORF14ng nucleotide sequence <SEQ ID 145> is predicted to encode a protein having amino acid sequence <SEQ ID 146>:

RSLTNPTVSVRIMLHAGLMYSRRAVVSRVAKSWSFAYMPDLVSRLNRLDLPTLV

GFFHGISVSSVFGAAAQYSAMASRSASIPVFSATEMRTAAIFPAASRHMPVFCSSDGSRS

```
1 MEDLQEIGFD VAAVKVGRQR EHHRLHHTQS GNGKADDVLF AFFLVGGFDF
51 LRVIGCGGVA CLPDFQQNVG EADFAVVPDD AAAVRAVIEV DADDAVCAQK
101 LLFDQPDAGG AGNAAEHQHC FVRAIMGFHK VGLDFGQVVQ ADLVEDFLGR
151 QFGFFRVGGA SFVITAQAGI DDALCDCLTA DAAGFAVFAF VADGQMQVFG
5 201 NVQPAVETGF FHGISVSSVF GAAAQYSAMA SRSASIPVFS ATEMRTAAIF
251 PAASRHMPVF CSSDGSRSVL LYTLMHGISW AWISCSTFST SSICCPLFRA
301 AASTTCSSTS ACTVSSKVAE KAEISLCGRS LTNPTVSVRI MLHAGLMYSR
351 RAVVSRVAKS WSFAYMPDLV SRLNRLDLPT LV*
```

Based on the putative transmembrane domain in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 18

The following partial DNA sequence was identified in N.meningitidis <SEQ ID 147>:

```
..GGCCATTACT CCGACCGCAC TTGGAAGCCG CGTTTGGNCG GCCGCCGTCT
15
                 51
                       GCCGTATCTG CTTTATGGCA CGCTGATTGC GGTTATTGTG ATGATTTTGA
                       TGCCGAACTC GGGCAGCTTC GGTTTCGGCT ATGCGTCGCT GGCGGCTTTG
                101
                151
                       TCGTTCGGCG CGCTGATGAT TGCGCTGTTA GACGTGTCGT CAAATATGGC
                       GATGCAGCCG TTTAAGATGA TGGTCGGCGA CATGGTCAAC GAGGAGCAGA
                201
                       AAA.NTACGC CTACGGGATT CAAAGTTTCT TAGCAAATAC GGGCGCGGTC
                251
                       GTGGCGGCGA TTCTGCCGTT TGTGTTTGCG TATATCGGTT TGGCGAACAC
20
                301
                       CGCCGANAAA GGCGTTGTGC CGCAGACCGT GGTCGTGGCG TTTTATGTGG
                351
                       GTGCGGCGTT GCTGGTGATT ACCAGCGCGT TCACGATTTT CAAAGTGAAG
                401
                       GAATACGANC CGGAAACCTA CGCCCGTTAC CACGGCATCG ATGTCGCCGC
                451
                       GAATCAGGAA AAAGCCAACT GGATCGCACT CTTAAAA.CC GCGC..
                501
```

25 This corresponds to the amino acid sequence <SEQ ID 148; ORF16>:

```
1 ..GHYSDRTWKP RLXGRRLPYL LYGTLIAVIV MILMPNSGSF GFGYASLAAL
51 SFGALMIALL DVSSNMAMQP FKMMVGDMVN EEQKXYAYGI QSFLANTGAV
101 VAAILPFVFA YIGLANTAXK GVVPQTVVVA FYVGAALLVI TSAFTIFKVK
151 EYXPETYARY HGIDVAANQE KANWIALLKX A..
```

30 Further work revealed the complete nucleotide sequence <SEQ ID 149>:

```
1 ATGTCGGAAT ATACGCCTCA AACAGCAAAA CAAGGTTTGC CCGCGCTGGC
                     AAAAAGCACG ATTTGGATGC TCAGTTTCGG CTTTCTCGGC GTTCAGACGG
                     CCTTTACCCT GCAAAGCTCG CAAATGAGCC GCATTTTCA AACGCTAGGC
                101
                     GCAGACCCGC ACAATTTGGG CTGGTTTTTC ATCCTGCCGC CGCTGGCGGG
                151
                     GATGCTGGTG CAGCCGATTG TCGGCCATTA CTCCGACCGC ACTTGGAAGC
35
                201
                     CGCGTTTGGG CGGCCGCCGT CTGCCGTATC TGCTTTATGG CACGCTGATT
                251
                     GCGGTTATTG TGATGATTTT GATGCCGAAC TCGGGCAGCT TCGGTTTCGG
                301
                     CTATGCGTCG CTGGCGGCTT TGTCGTTCGG CGCGCTGATG ATTGCGCTGT
                351
                     TAGACGTGTC GTCAAATATG GCGATGCAGC CGTTTAAGAT GATGGTCGGC
                401
                     GACATGGTCA ACGAGGAGCA GAAAGGCTAC GCCTACGGGA TTCAAAGTTT
40
                451
                     CTTAGCAAAT ACGGGCGCG TCGTGGCGGC GATTCTGCCG TTTGTGTTTG
                501
                     CGTATATCGG TTTGGCGAAC ACCGCCGAGA AAGGCGTTGT GCCGCAGACC
                551
                     GTGGTCGTGG CGTTTTATGT GGGTGCGGCG TTGCTGGTGA TTACCAGCGC
                601
                     GTTCACGATT TTCAAAGTGA AGGAATACGA TCCGGAAACC TACGCCCGTT
                651
                     ACCACGGCAT CGATGTCGCC GCGAATCAGG AAAAAGCCAA CTGGATCGAA
45
                701
                     CTCTTGAAAA CCGCGCCTAA GGCGTTTTGG ACGGTTACTT TGGTGCAATT
                751
                     CTTCTGCTGG TTCGCCTTCC AATATATGTG GACTTACTCG GCAGGCGCGA
                801
                     TTGCGGAAAA CGTCTGGCAC ACCACCGATG CGTCTTCCGT AGGTTATCAG
                851
                     GAGGCGGGTA ACTGGTACGG CGTTTTGGCG GCGGTGCAGT CGGTTGCGGC
                     GGTGATTTGT TCGTTTGTAT TGGCGAAAGT GCCGAATAAA TACCATAAGG
50
                951
                     CGGGTTATTT CGGCTGTTTG GCTTTGGGCG CGCTCGGCTT TTTCTCCGTT
               1001
                     TTCTTCATCG GCAACCAATA CGCGCTGGTG TTGTCTTATA CCTTAATCGG
               1051
                     CATCGCTTGG GCGGGCATTA TCACTTATCC GCTGACGATT GTGACCAACG
               1101
                     CCTTGTCGGG CAAGCATATG GGCACTTACT TGGGCTTGTT TAACGGCTCT
               1151
                     ATCTGTATGC CTCAAATCGT CGCTTCGCTG TTGAGTTTCG TGCTTTTCCC
 55
               1201
                     TATGCTGGGC GGCTTGCAGG CCACTATGTT CTTGGTAGGG GGCGTCGTCC
               1251
                     TGCTGCTGGG CGCGTTTTCC GTGTTCCTGA TTAAAGAAAC ACACGGCGGG
               1301
               1351
                     GTTTGA
```

This corresponds to the amino acid sequence <SEQ ID 150; ORF16-1>:

```
MSEYTPQTAK QGLPALAKST IWMLSFGFLG VQTAFTLQSS QMSRIFQTLG
                     51
                         ADPHNLGWFF ILPPLAGMLV QPIVGHYSDR TWKPRLGGRR LPYLLYGTLI
                   101
                          AVIVMILMPN SGSFGFGYAS LAALSFGALM IALLDVSSNM AMOPFKMMVG
                         DMVNEEQKGY AYGIQSFLAN TGAVVAAILP FVFAYIGLAN TAEKGVVPQT
VVVAFYVGAA LLVITSAFTI FKVKEYDPET YARYHGIDVA ANQEKANWIE
 5
                   151
                   201
                   251 LLKTAPKAFW TVTLVQFFCW FAFQYMWTYS AGAIAENVWH TTDASSVGYQ
                          EAGNWYGVLA AVQSVAAVIC SFVLAKVPNK YHKAGYFGCL ALGALGFFSV
                   301
                         FFIGNOYALV LSYTLIGIAW AGIITYPLTI VTNALSGKHM GTYLGLFNGS ICMPQIVASL LSFVLFPMLG GLOATMFLVG GVVLLLGAFS VFLIKETHGG
10
                   401
                   451
```

Computer analysis of this amino acid sequence gave the following results:

# Homology with a predicted ORF from N. meningitidis (strain A)

ORF16 shows 96.7% identity over a 181aa overlap with an ORF (ORF16a) from strain A of N.

### 15 meningitidis:

	orf16.pep			GH	10 YSDRTWKPRL	20 XGRRLPYLLY	30 GTLIAVIV
				71	1111111111	111111111	
	orf16a	IFQTLGADPHSI	GWFFILPPL	AGMLVQPIVGH	YSDRTWKPRL	GGRRLPYLLY	GTLIAVIV
20		50	60	70	80	90	100
•		40	50	60	7.0		
	orf16.pep				70	80	90
	Offig.beb	MILMPNSGSFG	GIASLAALS	GALMIALLDV	SSNMAMOPIK	MMVGDMVNE	EQKXYAYGI
25	orf16a	MILMONECCEC	CVACTARTO	11(11111111		111111111	
23	Offica	MILMPNSGSFG	GYASLAALS	GALMIALLDV			
		110	120	130	140	150	160
	•	100	110				
		100	110	120	130	140	. 150
20	orf16.pep	QSFLANTGAVV/	ALLPFVFAY.	<u>IGLA</u> NTAXKGV	VPQT <u>VVVAFY</u>	VGAALLVITS	AFTIFKVK
30	c1 c				111111111	11111111	3 1 1 1 1 1 1 1 1
	orf16a	QSFLANTGAVV					<u>A</u> FTIFKVK
	•	170	180	190	200	210	220
		160	170	180			
35	orf16.pep	EYXPETYARYHO					
	· OZZZZO · POP						
	orf16a	FYNDETYADYU	יוונוווווווווו משכת אל אל למלב		VA ESIMIEM I IIA	DECLIOS DATE	
	OLLIVA	EYNPETYARYHO 230	240	250			
		230	240	250	260	270	280
40	orf16a	A PARTALLANDO A CO	TICVOED CHEN	/CIII			
70	OLLIVA	AENVWHTTDASS 290	OVGIQEAGNW:	GVLAAVQSVA			
		290	300	310	320	330	340

# The complete length ORF16a nucleotide sequence <SEQ ID 151> is:

	1	ATGTCGGAAT	ATACGCCTCA	AACAGCAAAA	CAAGGTTTGC	CCGCGCTGGC
4.5	51				CTTTCTCGGC	
45	101	CCTTTACCCT	GCAAAGCTCG	CAGATGAGCC	GCATCTTCCA	GACGCTCGGT
	151	GCCGATCCGC	ACAGCCTCGG	CTGGTTCTTT	ATCCTGCCGC	CGCTGGCGGG
•	201	GATGCTGGTG	CAGCCGATTG	TCGGCCATTA	CTCCGACCGC	ACTTGGAAGC
	251			CTGCCGTATC		CACGCTGATT
	301	GCGGTTATTG	TGATGATTTT	GATGCCGAAC	TCGGGCAGCT	TCGGTTTCGG
50	351	CTATGCGTCG	CTGGCGGCTT	TGTCGTTCGG	CGCGCTGATG	ATTGCGCTGT
	401			GCGATGCAGC		GATGGTCGGC
	451	GACATGGTCA	ACGAGGAGCA	GAAAGGCTAC	GCCTACGGGA	TTCAAAGTTT
	501	CTTAGCGAAT	ACGGGCGCGG	TCGTGGCGGC	GATTCTGCCG	TTTGTGTTTG
	551	CGTATATCGG	TTTGGCGAAC	<b>ACCGCCGAGA</b>	AAGGCGTTGT	GCCGCAGACC
55	601	GTGGTCGTGG	CGTTTTATGT	GGGTGCGGCG	TTGCTGGTGA	TTACCAGCGC
	651	GTTCACGATT	TTCAAAGTGA	AGGAATACAA	TCCGGAAACC	TACGCCCGTT
	701	ACCACGGCAT	CGATGTCGCC	GCGAATCAGG	AAAAAGCCAA	CTGGATCGAA
	751	CTCTTGAAAA	CCGCGCCTAA	GGCGTTTTGG	ACGGTTACTT	TGGTGCAATT
	801	CTTCTGCTGG	TTCGCCTTCC	<b>AATATATGTG</b>	GACTTACTCG	GCAGGCGCGA
60	851	TTGCGGAAAA	CGTCTGGCAC	ACCACCGATG	CGTCTTCCGT	AGGTTATCAG
	901	GAGGCGGGTA	ACTGGTACGG	CGTTTTGGCG	GCGGTGCAGT	CGGTTGCGGC
•	951				GCCGAATAAA	

5	1051 TTCTT 1101 CATCG 1151 CCTTG 1201 ATCTG 1251 TATGG 1301 TGCTG 1351 GTTTG		CGCGCTGGT TCACTTATC GGCACTTAC CGCTTCGCT CCACTATGT GTGTTCCTG	G TTGTCTTA: C GCTGACGA: T TGGGCCTG: G TTGAGTTTC T CTTGGTAGCA A TTAAAGAA	FA CCTTAAT TT GTGACCA TT TAACGGC CG TGCTTTT GG GGCGTCG AC ACACGGC	CGG NACG CTCT CCC STCC	
	This encodes a protein						
10	51 ANDH	PPOTAK QGLPALAKST SLGWFF ILPPLAGMLV MILMPN SGSFGFGYAS	OPIVGHYSD	R TWKPRLGG	RR LPYLLY	<u> STLI</u>	
15	151 DMVNI 201 VVVA 251 LLKT. 301 EAGN 351 FFIG 401 ICMP 451 V*	EEQKGY AYGIQSFLAN FYVGAA LLVITSAFTI APKAFW TVTLVQFFCW WYGVLA AVQSVAAVIO NOYALV LSYTLIGIAW QIVASL LSFVLFPMLG	TGAVVAAIL FKVKEYNPE FAFQYMWTY SFVLAKVPN AGIITYPLT GLQATMFLV	P FVFAYIGL T YARYHGID S AGAIAENV K YHKAGYFG I VTNALSGK G GVVLLLGA	AN TAEKGVV VA ANQEKAN WH TTDASSV CL ALGALGI HM GTYLGLI	VPQT NWIE VGYQ FFSV FNGS	
20	ORF16a and ORF16-	1 show 99.6% iden	tity in 451	aa overlap:			
	orf16a.pep orf16-1	10 MSEYTPQTAKQGLPAI			11111111	11111111	:     {
25		10			100	110	120
	orf16a.pep	70 	111111111		GTLIAVIVM	ILMPNSGSF0	GFGYAS
30	orf16-1	ILPPLAGMLVQPIVG	HYSDRTWKPR 80	LGGRRLPYLLY 90	GTLIAVIVM 100	ILMPNSGSF	GFGYAS 120
35	orf16a.pep orf16-1	130 LAALSFGALMIALLD             LAALSFGALMIALLD	HILLIH HILL	1111111111	1111111111		11111
	01110-1	130	140	150	160	170	180
40	orf16a.pep	190 FVFAYIGLANTAEKG	1111111111	111111111	111111111	:	111111
	orf16-1	FVFAYIGLANTAEKG 190	VVPQTVVVAF 200	YVGAALLVIT 210	SAFTIFKVKE 220	EYDPETYARY 230	HGIDVA 240
45	orf16a.pep	250 ANQEKANWIELLKTA	260 PKAFWTVTLV	270 QFFCWFAFQY	280 MWTYSAGAI	290 AENVWHTTDA	300 SSVGYQ
	orf16-1					11111111111111111111111111111111111111	ASSVGYQ 300
50		310	320	330	340	350	360
	orfl6a.pep	EAGNWYGVLAAVQSV			111111111	11 [11 ] [11 ]	
55	orf16-1	EAGNWYGVLAAVQSV 310	AAVICSFVLA 320	330	340	350	360
	orfl6a.pep	370 LSYTLIGIAWAGII	380 CYPLTIVTNA	390 LSGKHMGTYL	400 SLFNGSICMP	410 QIVASLLSEV	420 VLFPMLG
60	orf16-1				GLFNGSICMP 400	QIVASLLSE 410	VLFPMLG 420
65	orf16a.pep orf16-1	430 GLQATMFLVGGVVL              GLQATMFLVGGVVL 430	1111111111	1111111			

# Homology with a predicted ORF from N.gonorrhoeae

ORF16 shows 93.9% identity over a 181aa overlap with a predicted ORF (ORF16.ng) from N. gonorrhoeae:

	gonorrhoeae:		
5	orf16.pep	GHYSDRTWKPRLXGRRLPYLLYGTLIAVIV	30
	orfl6ng	HFSNARRRPAQFGLVFHPAAAGGDAGSADSGYYSDRTWKPRLGGRRLPYLLYGTLIAVIV	131
10	orf16.pep	MILMPNSGSFGFGYASLAALSFGALMIALLDVSSNMAMQPFKMMVGDMVNEEQKXYAYGI	90
	orf16ng	MILMPNSGSFGFGYASLAALSFGALMIALLDVSSNMAMQPFKMMVGDMVNEEQKSYAYGI	191
	orf16.pep	QSFLANTGAVVAAILPFVFAYIGLANTAXKGVVPQTVVVAFYVGAALLVITSAFTIFKVK	150
15	orf16ng	QSFLANTDAVVAAILPFVFAYIGLANTAEKGVVPQTVVVAFYVGAALLIITSAFTISKVK	251
	orf16.pep	EYXPETYARYHGIDVAANQEKANWIALLKXA	181
	orf16ng		311
20	The complete leng	gth ORF16ng nucleotide sequence <seq 153="" id=""> is:</seq>	
-	51 T 101 C	TGATAGGGG ATCGCCGCGC CGGCAACCAT TTCGGATTTT CCAAAGCAAA CACTTTTCAA ATCAAAAAAA AGGATTTACT TTATGTCGGA ATATACGCCT CAAACAGCAA AACAAGGTTT GCCCGGCCG GCAAAAAGCA CGATTTGGAT	
25	201 0 251 0 301 F	TTTGAGCTTC GGCTATCTCG GCGTTCAGAC GGCCTTTACC CTGCAAAGCT CGCAGATGAG CCGCATTTTT CAAACGCTAG GCGCAGACCC GCACAATTTG GGCTGGTTTT TCATCCTGCC GCCGCTGGCG GGGATGCTGG TTCAGCCGAT AGTGGCTACT ACTCAGACCG CACTTGGAAG CCGCGCTTGG GCGGCCGCCG CCTGCCGTAT CTGCTTTACG GCACGCTGAT TGCGGTCATC GTGATGATTT	
30	401 1 451 1 501 6 551 7	CATGCCGAA CTCGGGCAGC TTCGGTTTCG GCTATGCGTC GCTGGCGGCC CTGTCGTTCG GCGCCTGAT GATTGCGCTG TTGGACGTGT CGTCGAATAT GGCGATGCAG CCGTTTAAGA TGATGGTCGG CGATATGGTC AACGAGGAGC AGAAAAGCTA CGCCTACGGG ATTCAAAGTT TCTTAGCGAA TACGGACGCG	٠.
35	651 C 701 T 751 F	STTGTGGCAG CGATTCTGCC GTTTGTGTTC GCGTATATCG GTTTGGCGAA CACTGCCGAG AAAGGCGTTG TGCCACAAAC CGTGGTCGTA GCATTCTATG CGGTGCGGC GTTACTGATT ATTACCAGTG CGTTCACAAT CTCCAAAGTC LAAGAATACG ACCCGGAAAC CTACGCCCGT TACCACGGCA TCGATGTCGC CGCGAATCAG GAAAAAGCCA ACTGGTTCGA ACTCTTAAAA ACCGCGCCTA	
40	851 # 901 C 951 C	AGTGTTTTG GACGGTTACT CCGGTACAGT TTTTCTGCTG GTTCGCCTTC CGGTATATGT GGACTTACTC GGCAGGCGCG ATTGCAGAAA ACGTCTGGCA CACTACCGAT GCGTCTTCCG TAGGCCATCA GGAGGCGGGC AACCGGTACG CCGTTTTGGC GGCGGTGTAG	÷
	This encodes a pro	otein having amino acid sequence <seq 154="" id="">:</seq>	
45	51 V 101 S 151 I 201 V 251 F	MIGDRRAGNH FGFSKANTFQ IKKKDLLYVG IYASNSKTRF ARAGKKHDLD VELRLSRRSD GLYPAKLADE PHFSNARRRP AQFGLVFHPA AAGGDAGSAD GGYYSDRTWK PRLGGRRLPY LLYGTLIAVI VMILMPNSGS FGFGYASLAA LSFGALMIAL LDVSSNMAMQ PFKMMVGDMV NEEQKSYAYG IQSFLANTDA VVAAILPFVF AYIGLANTAE KGVVPQTVVV AFYVGAALLI ITSAFTISKV GYDPETYAR YHGIDVAANQ EKANWFELLK TAPKVFWTVT PVQFFCWFAF RYMWTYSAGA IAENVWHTTD ASSVGHQEAG NRYGVLAAV*	
50	ORF16ng and OR	F16-1 show 89.3% identity in 261 aa overlap:	
	orf16-1.pep	30 40 50 60 70 80  MLSFGFLGVQTAFTLQSSQMSRIFQTLGADPHNLGWFFILPPLAGMLVQPI-VGHYSDI	RT
55	orf16ng	DVELRLSRRSDGLYPAKLADEPHFSNARRRPAQFGLVF-HPAAAGGDAGSADSGYYSDF 50 60 70 80 90 100	RT
60	orf16-1.pep	90 100 110 120 130 140  WKPRLGGRRLPYLLYGTLIAVIVMILMPNSGSFGFGYASLAALSFGALMIALLDVSSNI	4A.
		110 120 130 140 150 160	

5	orf16-1.pep	  MQPFKMMVG		170 YGIQSFLANTGA             YGIQSFLANTDA  90 200	VVAAILPEVEA	111111111	11111
10	orf16-1.pep	 VVAFYVGAA	LLVITSAFTII   :        LLIITSAFTIS	230 FKVKEYDPETYAI            SKVKEYDPETYAI 250 260		1111:11111	11:11
15	orf16-1.pep	11 111111	FAFQYMWTYS    :       FAFRYMWTYS	290 AGAIAENVWHTT           AGAIAENVWHTT 310 32	:      DASSVGHQEAGN		320 MAAVICS

Based on this analysis, including the presence of several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 19

The following partial DNA sequence was identified in N.meningitidis <SEQ ID 155>:

```
25

1 ATGTTGTTCC GTAAAACGAC CGCCGCCGTT TTGGCGCATA CCTTGATGCT
51 GAACGGCTGT ACGTTGATGT TGTGGGGAAT GAACAACCCG GTCAGCGAAA
101 CAATCACCCG NAAACACGTT GNCAAAGACC AAATCCGNGN CTTCGGTGTG
151 GTTGCCGAAG ACAATGCCCA ATTGGAAAAG GGCAGCCTGG TGATGATGGG
201 CGGAAAATAC TGGTTCGTCG TCAATCCCGA AGATTCGGCG AA.NTGACGG
301 CCGAGCTATG C.TGCCACCA AGCCCTGCCG GTCAAACTCC GATCGNCTGG
351 CAGCCAGAAT...
```

This corresponds to the amino acid sequence <SEQ ID 156; ORF28>:

```
35 MLFRKTTAAV LAHTIMINGC TLMLWGMNNP VSETITRKHV XKDQIRXFGV
51 VAEDNAQIEK GSLVMMGGKY WFVVNPEDSA XXTGILXAGL DKPFQIVXDT
101 PSYXCHQALP VKLGSXGSQN...
```

Further work revealed the complete nucleotide sequence <SEQ ID 157>:

1 ATGTTGTTCC GTAAAACGAC CGCCGCCGTT TTGGCGGCAA CCTTGA 51 GAACGGCTGT ACGTTGATGT TGTGGGGAAT GAACAACCCG GTCAGC	GAAA
51 GAACGGCTGT ACGTTGATGT TGTGGGGAAT GAACAACCCG GTCAGC	
40 101 CARTCACCG CARACAGGTT GACAAAGACC AAATCCGCGC CTTCGG	TGTG
151 CTTCCCGAAG ACAATGCCCA ATTGGAAAAG GGCAGCCTGG TGATGA	TGGG
201 CCCDDATAC TGGTTCGTCG TCAATCCCGA AGATTCGGCG AAGCTG	ACGG
251 CCATTTTGAA GGCAGGGCTG GACAAACCCT TCCAAATAGT TGAGGA	TACC
301 CCGAGCTATG CTCGCCACCA AGCCCTGCCG GTCAAACTCG AATCGC	CTGG
AS 351 CAGCCAGAAT TTCAGTACCG AAGGCCTTTG CCTGCGCTAC GATACC	GACA
A01 AGCCTGCCGA CATCGCCAAG CTGAAACAGC TCGGGTTTGA AGCGGT	CAAA
451 CTCGACAATC GGACCATTTA CACGCGCTGC GTATCCGCCA AAGGCA	ATAA
501 CTACGCCACA CCGCAAAAAC TGAACGCCGA TTACCATTTT GAGCAA	AGTG
551 TCCCTCCCA TATTTATTAC ACGGTTACTG AAGAACATAC CGACAA	ATCC
50 601 AAGCTGTTTG CAAATATCTT ATATACGCCC CCCTTTTTGA TACTGG	ATGC
651 GGCGGGCGC GTACTGGCCT TGCCTGCGGC GGCTCTGGGT GCGGTC	:GTGG
701 ATGCCGCCCG CAAATGA	

This corresponds to the amino acid sequence <SEQ ID 158; ORF28-1>:

	1	MLFRKTTAAV	LAATLMLNGC	TLMLWGMNNP	VSETITRKHV	DKDQIRAFGV
55	51	VAEDNAQLEK	GSLVMMGGKY	WEVVNPEDSA	KLTGILKAGL	DKPFQIVEDT
55	101	PSYARHQALP	VKLESPGSQN	FSTEGLCLRY	DTDKPADIAK	LKQLGFEAVK
	151	LDNRTIYTRC	VSAKGKYYAT	PQKLNADYHF	EQSVPADIYY	TVTEEHTDKS

# 201 KLFANILYTP PFLILDAAGA VLALPAAALG AVVDAARK\*

Computer analysis of this amino acid sequence gave the following results:

# Homology with a predicted ORF from N.meningitidis (strain A)

ORF28 shows 79.2% identity over a 120aa overlap with an ORF (ORF28a) from strain A of N.

### 5 meningitidis:

		10			. 40	50	60
	orf28.pep	MLFRKTTAAV	/LAHTLMLNGC	TLMLWGMNNP	<b>VSETITRKHV</b>	KKDQIRXFGV	/AEDNAQLEK
		- 11111111111		1:1:1111:1		11111 1111	1111111111
10	orf28a		/LAATLMLNGC		FSETTARKHVI	OKDQIRAFGV	/AEDNAQLEK
10	•	10	20	30	40	50	60
		7.					
		70	-		100	110	120
	orf28.pep	GSLVMMGGKY	wfvvnpedsa	XXTGILXAGL	DKPFQIVXDTI	SYXCHQALP	/KLGSXGSON
			[[[]]]	1111 111		: : : : : : : : : : : : : : : : : : : :	111:1:11
15	orf28a	GSLVMMGGK	/wfvvnpedsa	KLTGILKAGL	DKQFQMVEPNI	PRFA-YOALP	KLESPASON
		70	80		100	110	
	orf28a	FSTEGLCLR1	/DTDRPADIAK	TROTE PERMITE	T DAIDMTVMD CT	7C N 1/C 1/C1/23/ N m v	
	011200	120 13					
		120 13	30 14	0 15	0 160	170	)

# 20 The complete length ORF28a nucleotide sequence <SEQ ID 159> is:

•						
	. 1	ATGTTGTTCC	GTAAAACGAC	CGCCGCCGTT	TTGGCGGCAA	CCTTGATGTT
	51				GAACAGCCCG	
	101				AAATCCGCGC	
	151	GTTGCCGAAG	ACAATGCCCA	ATTGGAAAAG	GGCAGCCTGG	TGATGATGGG
25	201	CGGGAAATAC	TGGTTCGTCG	TCAATCCTGA	AGATTCGGCG	AAGCTGACGG
	251	GCATTTTGAA	GGCCGGGTTG	GACAAGCAGT	TTCAAATGGT	TGAGCCCAAC
	301	CCGCGCTTTG	CCTACCAAGC	CCTGCCGGTC	AAACTCGAAT	CGCCCGCCAG
	351	CCAGAATTTC	AGTACCGAAG	GCCTTTGCCT	GCGCTACGAT	ACCGACAGAC
	401	CTGCCGACAT	CGCCAAGCTG	AAACAGCTTG	AGTTTGAAGC	GGTCGAACTC
30	451	GACAATCGGA	CCATTTACAC	GCGCTGCGTC	TCCGCCAAAG	GCAAATACTA
	501	CGCCACACCG	CAAAAACTGA	<b>ACGCCGATTA</b>	TCATTTTGAG	CAAAGTGTGC
	551	CTGCCGATAT	TTATTACACG	GTTACGAAAA	<b>AACATACCGA</b>	CAAATCCAAG
	601				ACGTTGATAC	
	651	GGGCGCGGTG	CTGGCCTTGC	CTGTCGCGGC	GTTGATTGCA	GCCACGAATT
35	701	CCTCAGACAA				

This encodes a protein having amino acid sequence <SEQ ID 160>:

	1	MLFRKTTAAV	LAATLMLNGC	TVMMWGMNSP	<b>FSETTARKHV</b>	DKDOIRAFGV
					KLTGILKAGL	
	101	PRFAYQALPV	KLESPASONF	STEGLCLRYD	TDRPADIAKL	KOLEFEAVEL
40	151	DNRTIYTRCV	SAKGKYYATP	QKLNADYHFE	QSVPADIYYT	VTKKHTDKSK
	201	LFENIAYTPT	TLILDAVGAV	LALPVAALIA	ATNSSDK*	

### ORF28a and ORF28-1 show 86.1% identity in 238 aa overlap:

	orf28a.pep	MLFRK	10 FTAAVLAAT	20 LMLNGCTVMM	30 WGMNSPFSE	40 FTARKHVDKDQ	50 IRAFGVVAED	60 NAOLEK
45	orf28-1	1111	  TAAVLAAT	: : LMLNGCTLML	1111:1 11		1111111111	HĪHE
			10	20	30	40	50	. 60
50			70	80	90	100	110	119
50	orf28a.pep	GSLVM	MGGKYWFVVI	NPEDSAKLTG 	ILKAGLDKQ:	FOMVEPNPRFA		
	orf28-1	GSLVM	MGGKYWFVVI	NPEDSAKLTG	ILKAGLDKP:	:   :  :  FQIVEDTPSYA	:      : RHQALPVKLE	II:III SPGSON
			70.	80	90	100	110	120
55		120	130	140	150	160	170	179
	orf28a.pep	FSTEG	LCLRYDTDR	PADIAKLKOL		RTIYTRCVSAK	GKYYATPQKL	NADYHF
	orf28-1		LCLRYDTDK	PADIAKLKQL				 NADYHF
			130	140	150	160	170	100

5	orf28a.pep orf28-1	180 190 200 210 220 230  EQSVPADIYYTVTKKHTDKSKLFENIAYTPTTLILDAVGAVLALPVAALIAATNSSDKX
		edicted ORF from N.gonorrhoeae
	ORF28 shows 84.2%	6 identity over a 120aa overlap with a predicted ORF (ORF28.ng) from N.
10	gonorrhoeae:	
	orf28.pep	MLFRKTTAAVLAHTLMLNGCTLMLWGMNNPVSETITRKHVXKDQIRXFGVVAEDNAQLEK 60
	orf28ng	MLFRKTTAAVLAATLILNGCTMMLRGMNNPVSQTITRKHVDKDQIRAFGVVAEDNAQLEK 60
15		GSLVMMGGKYWFVVNPEDSAXXTGILXAGLDKPFQIVXDTPSYXCHQALPVKLGSXGSQN 120
	orf28ng (	GSLVMMGGKYWFAVNPEDSAKLTGLLKAGLDKPFQIVEDTPSYARHQALPVKFEAPGSQN 120
	<u>-</u>	ORF28ng nucleotide sequence <seq 161="" id=""> is</seq>
20	51 GAA 101 CAA 151 GTT	TTGTTCC GTAAAACGAC CGCCGCCGTT TTGGCGGCAA CCTTGATACT CGGCTGT ACGATGATGT TGCGGGGGAT GAACAACCCG GTCAGCCAAA ICACCCG CAAACACGTT GACAAAGACC AAATCCGCGC CTTCGGTGTG GCCGAAG ACAATGCCCA ATTGGAAAAG GGCAGCCTGG TGATGATGGG
25	251 GCC: 301 CCGA 351 CAGC 401 GACC 451 CTCC	GAAATAC TGGTTCGCCG TCAATCCCGA AGATTCGGCG AAGCTGACGG ITTTGAA GGCCGGGTTG GACAAGCCCT TCCAAATAGT TGAGGATACC AGCTATG CCCGCCACCA AGCCCTGCCG GTCAAATTCG AAGCGCCCGG CCAGAAT TTCAGTACCG GAGGTCTTTG CCTGCGCTAT GATACCGGCA CTGACGA CATCGCCAAG CTGAAACAGC TTGAGTTTAA AGCGGTCAAA GACAATC GGACCATTTA CACGCGCTGC GTATCCGCCA AAGGCAAATA CGCCACG CCGCAAAAAC TGAACGCCGA TTATCATTTT GAGCAAAGTG
30	551 TGC0 601 AAG0 651 GGC0	CCGCCGA TATTATTAT ACGGTTACTG AAAAACATAC CGACAAATCC CTGTTTG GAAATATCTT ATATACGCCC CCCTTGTTGA TATTGGATGC GGCCGCG GTGCTGGTCT TGCCTATGGC TCTGATTGCA GCCGCGAATT CAGACAA ATGA
	This encodes a protein	in having amino acid sequence <seq 162="" id="">:</seq>
35	51 VAEI 101 PSYX 151 LDNI	RKTTAAV LAATLILNGC TMMLRGMNNP VSQTITRKHV DKDQIRAFGV DNAQLEK GSLVMMGGKY WFAVNPEDSA KLTGLLKAGL DKPFQIVEDT ARHQALP VKFEAPGSQN FSTGGLCLRY DTGRPDDIAK LKQLEFKAVK RTIYTRC VSAKGKYYAT PQKLNADYHF EQSVPADIYY TVTEKHTDKS GNILYTP PLLILDAAAA VLVLPMALIA AANSSDK*
40	ORF28ng and ORF2	8-1 share 90.0% identity in 231 aa overlap:
	orf28-1.pep	10 20 30 40 50 60 MLFRKTTAAVLAATLMLNGCTLMLWGMNNPVSETITRKHVDKDQIRAFGVVAEDNAQLEK
45	orf28ng	MLFRKTTAAVLAATLILNGCTMMLRGMNNPVSQTITRKHVDKDQIRAFGVVAEDNAQLEK 10 20 30 40 50 60
50	orf28-1.pep	70 80 90 100 110 120 GSLVMMGGKYWFVVNPEDSAKLTGILKAGLDKPFQIVEDTPSYARHQALPVKLESPGSQN
50	orf28ng	GSLVMMGGKYWFAVNPEDSAKLTGLLKAGLDKPFQIVEDTPSYARHQALPVKFEAPGSQN 70 80 90 100 110 120
55	orf28-1.pep orf28ng	130 140 150 160 170 180  FSTEGLCLRYDTDKPADIAKLKQLGFEAVKLDNRTIYTRCVSAKGKYYATPQKLNADYHF            :             :
60	orf28-1.pep orf28ng	190 200 210 220 230 239 EQSVPADIYYTVTEEHTDKSKLFANILYTPPFLILDAAGAVLALPAAALGAVVDAARKX

PCT/IB98/01665 WO 99/24578 -143-

200

190

210

220

230

Based on this analysis, including the presence of a putative transmembrane domain in the gonococcal protein, it was predicted that the proteins from N. meningitidis and N. gonorrhoeae, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF28-1 (24kDa) was cloned in pET and pGex vectors and expressed in E.coli, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 6A shows the results of affinity purification of the GST-fusion protein, and Figure 6B shows the results of expression of the His-fusion in E.coli. Purified GST-fusion protein was used to immunise mice, whose sera were used for ELISA, which gave a positive result. These experiments confirm that ORF28-1 is a surface-exposed protein, and that it may be a useful immunogen.

#### Example 20

5

10

25

The following partial DNA sequence was identified in N. meningitidis <SEO ID 163>:

```
..GTCAGTCCTG TACTGCCTAT TACACACGAA CGGACAGGGT TTGAAGGTGT
15
                51
                       TATCGGTTAT GAAACCCATT TTTCAGGGCA CGGACATGAA GTACACAGTC
                101
                       CGTTCGATCA TCATGATTCA AAAAGCACTT CTGATTTCAG CGGCGGTGTA
                151
                       GACGGCGGTT TTACTGTTTA CCAACTTCAT CGAACATGGT CGGAAATCCA
                       TCCGGAGGAT GAATATGACG GGCCGCAAGC AGCG.ATTAT CCGCCCCCG
                201
                251
                       GAGGAGCAAG GGATATATAC AGCTATTATG TCAAAGGAAC TTCAACAAAA
20
                301
                       ACAAAGACTA GTATTGTCCC TCAAGCCCCA TTTTCAGACC GTTGGCTAGA
                       AGAAAATGCC GGTGCCGCCT CTGGT..
```

This corresponds to the amino acid sequence <SEQ ID 164; ORF29>:

```
.. VSPVLPITHE RTGFEGVIGY ETHFSGHGHE VHSPFDHHDS KSTSDFSGGV
       DGGFTVYQLH RTWSEIHPED EYDGPQAAXY PPPGGARDIY SYYVKGTSTK
101
       TKTSIVPQAP FSDRWLEENA GAASG..
```

Further work revealed the complete nucleotide sequence <SEQ ID 165>:

	1	ATGAATTTGC	CTATTCAAAA	ATTCATGATG	CTGTTTGCAG	CAGCAATATC
	51	GTTGCTGCAA	ATCCCCATTA	GTCATGCGAA	CGGTTTGGAT	GCCCGTTTGC
	101	GCGATGATAT	GCAGGCAAAA	CACTACGAAC	CGGGTGGTAA	ATACCATCTG
30	151	TTTGGTAATG	CTCGCGGCAG	TGTTAAAAAG	CGGGTTTACG	CCGTCCAGAC
	201	ATTTGATGCA	ACTGCGGTCA	GTCCTGTACT	GCCTATTACA	CACGAACGGA
	251	CAGGGTTTGA	AGGTGTTATC	GGTTATGAAA	CCCATTTTTC	AGGGCACGGA
	301	CATGAAGTAC	ACAGTCCGTT	CGATCATCAT	GATTCAAAAA	GCACTTCTGA
	351	TTTCAGCGGC	GGTGTAGACG	GCGGTTTTAC	TGTTTACCAA	CTTCATCGAA
35	401	CAGGGTCGGA	AATCCATCCG	GAGGATGGAT	ATGACGGGCC	GCAAGGCAGC
	451	GATTATCCGC	CCCCGGAGG	AGCAAGGGAT	ATATACAGCT	ATTATGTCAA
•	501	AGGAACTTCA	ACAAAAACAA	<b>AGACTAATAT</b>	TGTCCCTCAA	GCCCCATTTT
	551	CAGACCGTTG	GCTAAAAGAA	AATGCCGGTG	CCGCCTCTGG	TTTTTTCAGC
	601	CGTGCGGATG	AAGCAGGAAA	ACTGATATGG	GAAAGCGACC	CCAATAAAAA
40	651	TTGGTGGGCT	AACCGTATGG	ATGATGTTCG	CGGCATCGTC	CAAGGTGCGG
, -	.701	TTAATCCTTT	TTTAATGGGT	TTTCAAGGAG	TAGGGATTGG	GGCAATTACA
	751	GACAGTGCAG	TAAGCCCGGT	CACAGATACA	GCCGCGCAGC	AGACTCTACA
	801	AGGTATTAAT	GATTTAGGAA	AATTAAGTCC	GGAAGCACAA	CTTGCTGCCG
	851	CGAGCCTATT	ACAGGACAGT	GCTTTTGCGG	TAAAAGACGG	TATCAACTCT
45	901	GCCAAACAAT	GGGCTGATGC	CCATCCAAAT	<b>ATAACAGCTA</b>	CTGCCCAAAC
	951	TGCCCTTTCC	GCAGCAGAGG	CCGCAGGTAC	GGTTTGGAGA	GGTAAAAAAG
•	1001	TAGAACTTAA	CCCGACTAAA	TGGGATTGGG	TTAAAAATAC	CGGTTATAAA
	1051	AAACCTGCTG	CCCGCCATAT	GCAGACTTTA	GATGGGGAGA	TGGCAGGTGG
	1101	GAATAAACCT	ATTAAATCTT	TACCAAACAG	TGCCGCTGAA	AAAAGAAAAC
50	1151	AAAATTTTGA	GAAGTTTAAT	AGTAACTGGA	GTTCAGCAAG	TTTTGATTCA
<del>-</del> -						

PCT/IB98/01665 WO 99/24578 -144-

```
1201 GTGCACAAAA CACTAACTCC CAATGCACCT GGTATTTTAA GTCCTGATAA
                        AGTTAAAACT CGATACACTA GTTTAGATGG AAAAATTACA ATTATAAAAG
                        ATAACGAAAA CAACTATTTT AGAATCCATG ATAATTCACG AAAACAGTAT CTTGATTCAA ATGGTAATGC TGTGAAAACC GGTAATTTAC AAGGTAAGCA
                 1351
5
                        AGCAAAAGAT TATTTACAAC AACAAACTCA TATCAGGAAC TTAGACAAAT
                 1451 GA
```

This corresponds to the amino acid sequence <SEQ ID 166; ORF29-1>:

```
MNLPIQKFMM LFAAAISLLQ IPISHANGLD ARLRDDMQAK HYEPGGKYHL
                       FGNARGSVKK RVYAVQTFDA TAVSPVLPIT HERTGFEGVI GYETHFSGHG
10
                  101
                       HEVHSPFDHH DSKSTSDFSG GVDGGFTVYQ LHRTGSEIHP EDGYDGPQGS
                       DYPPPGGARD IYSYYVKGTS TKTKTNIVPQ APFSDRWLKE NAGAASGFFS
                  201 RADEAGKLIW ESDPNKNWWA NRMDDVRGIV QGAVNPFLMG FQGVGIGAIT
                  251 DSAVSPVTDT AAQQTLQGIN DLGKLSPEAQ LAAASLLQDS AFAVKDGINS
                  301 AKQWADAHPN ITATAQTALS AAEAAGTVWR GKKVELNPTK WDWVKNTGYK
15
                       KPAARHMOTL DGEMAGGNKP IKSLPNSAAE KRKONFEKFN SNWSSASFDS
VHKTLTPNAP GILSPDKVKT RYTSLDGKIT IIKDNENNYF RIHDNSRKQY
                  401
                  451 LDSNGNAVKT GNLQGKQAKD YLQQQTHIRN LDK*
```

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from N. meningitidis (strain A)

ORF29 shows 88.0% identity over a 125aa overlap with an ORF (ORF29a) from strain A of N. 20 meningitidis:

	orf29.pep			ven	10	20	30
	Offiza.pep					GFEGVIGYET     :	
25	orf29a	EPGGKYHLFGNAR	GSVKNRVYA				
		50	60	70	80	90	100
		40	50	60	70	80	90
	orf29.pep	VHSPFDHHDSKST	SDFSGGVDG	FTVYQLHRTW	SEIHPEDEY	DGPQAAXYPP	PGGARDIY
30					1111111		1111111
	orf29a	VHSPFDNHDSKST			SEIHPEDGY	DGPQGSDYPP	PGGARDIY
		110	120	130	140	150	160
		100	110	120			
35	orf29.pep	SYYVKGTSTKTKT	SIVPQAPFSI	ORWLEENAGAA	.SG		
		1111111111:					
	orf29a	XXYVKGTSTKTKS			SGFFSRADE	AGKLIWESDP:	NKNWWANR
		170	180	190	200	210	220
40	orf29a	MDDIRGIVQGAVN	PFLMGFOGV	GIGAITDSAVS	PVTDTAA00'	TLOGXNH1.GX	LSPEAOLA
		230	240	250	260	270	280

The complete length ORF29a nucleotide sequence <SEO ID 167> is:

	1	ATGAATTNGC	CTATTCAAAA	ATTCATGATG	CTGTTTGCAG	CAGCAATATC
	51	GTNGCTGCAA	ATCCCNATTA	GTCATGCGAA	CGGTTTGGAT	GCCCGTTTGC
45	101	GCGATGATAT	GCAGGCAAAA	CACTACGAAC	CGGGTGGTAA	ATACCATCTG
	151	TTTGGTAATG	CTCGCGGCAG	TGTTAAAAAT	CGGGTTTACG	CCGTCCAAAC
	201	ATTTGATGCA	ACTGCGGTCG	GCCCCATACT	GCCTATTACA	CACGAACGGA
•	251	CAGGATTTGA	AGGCATTATC	GGTTATGAAA	CCCATTTTTC	AGGACATGGA
	301				GATTCAAAAA	
50	351				CGTTTACCAA	
	401				ATGACGGGCC	
	451				ATATACANNT	
	501				TGTTCCCCGA	
	551				CCGCCTCTGG	
55	601				GAAAGCGACC	
	651				CGGCATCGTC	
	701				TAGGGATTGG	
	751				GCCGCGCAGC	
<b>60</b>	801				CGAAGCACAA	
60	851				TAAAAGACGG	
	901	GCCAGACAAT	GGGCTGATGC	CCATCCGAAT	ATAACTGCAA	CAGCCCAAAC

		CTTGCC GTAGCAGANG					
		ACTTAA CCCGACCAAA CTGCTG TTCGCACCAT					
	1051 ACAC 1101 GAAT	AGACCG CCTAAATCTA	GCATACTTTG	GATGGGGAAA	A TGGCCGGT	GG	
5	1101 GAA1	ACCGTC TTTACAAGCG	CARCURA	CAGUAAAGUA	A GATGCTTC	CA	
,	1201 CATG	CTTATA ACAAGCATGT	CAMCIAMIIG	CARCAACAAAI	CCCARRE		
	1251 TATO	CAATTCA CCAGCAGATT	TTTCCTCCCCA	TAMOMATTIA	A CGGATTTA	AA ·	
	1301 ATC	CANCAAA TATGAAAGAG	TIGCICGGCA	CUNCANCEC	CONTROLLAG	CC DM	
		AAACAG GGACNATAGT					•
10		GCATTT AGACCAACAT					
10	,I401 Inor	OCHITI AGACCAACAI	CAGGIAAAA	MIMITAIGAI	GALITAIA	G	
	This encodes a protei	n having amino acid	l sequence <	SEQ ID 168	ß>:		
	1 MNXE	PIQKFMM LFAAAISXLQ	TOTSHANCIN	ADT DODMONE	, naccen	tu T	
	51 FGNA	RGSVKN RVYAVQTFDA	TAVCOLL	REDUCERCE	CAEGRESC	nr nr	
		ISPFDNH DSKSTSDFSG					
15		PGGARD IYXXYVKGTS	TKTKSNIVPR	APESDRULKE	EDGIDGEQ	re re	
	201 RADE	AGKLIW ESDPNKNWWA	NRMDDIRGIV	OGAVNPFIMO	FOGVETER	TΨ	
	251 DSAV	SPVTDT AAQQTLQGXN	HLGXLSPEAO	LAAATALODS	S AFAVKOGT	NS	
		ADAHPN ITATAQTALA	VAXAATTVWG	GKKVELNPTE	WDWVKNTG	YX	
	351 TPAV	RTMHTL DGEMAGGNRP	PKSITSNSKA	DASTOPSLOA	OLIGEOTX	XG	
20	401 HAYN	KHVIRQ QEFTDLNINS	PADFARHIEN	IVSHPXNMKE	LPRGRTAY	WD	
	451 XKTG	TIVIRD KNSDDGGTAF	RPTSGKKYYD	DL*			
	ORF29a and ORF29-	1 show 90.1% iden	tity in 385 aa	a overlap:			
			-	•			
•	500	10	20	30	40	50 60	0
25	orf29a.pep	MNXPIQKFMMLFAAAI	SXLQIPISHAN	GLDARLRDDMC	<b>DAKHYEPGGK</b>	YHLFGNARGSVKI	N
23	orf29-1	MIL DECKENAL ENDER				!!!!!!!!!!!	:
	01129-1	MNLPIQKFMMLFAAAI					
		10	20	30	40	50 60	0
		70	80	90 1	100	110 120	^
30	orf29a.pep	RVYAVQTFDATAVGPI		CTTCVETHESC	CHCHEAMG DE LOO	UNINDERGAGDES	~
JU .	OIIIJu.pop		111111111111	1:11111111	HILLILL	1.1111111111	ۍ. ا
	orf29-1	RVYAVQTFDATAVSPV	LPITHERTGFE	GVIGYETHESO	CHCHEVHSPE	DHHDSKSTSDFS(	1.
		70	80			110 120	
	•						•,
35		130	140	150 1	L60	170 180	0
	orf29a.pep	GVDGGFTVYQLHRTGS	EIHPEDGYDGP	QGSDYPPPGG#	ARDIYXXYVK	GTSTKTKSNIVP	R
				11111111111	1111 1111	1111111:1111:	:
	.orf29-1	GVDGGFTVYQLHRTGS		QGSDYPPPGG#	ARDIYSYYVK	GTSTKTKTNIVP(	2
40		130	140	150 1	L60	170 180	0
40							
	orf29a.pep	190		210 2	220	230 240	0
	orizaa.pep	APFSDRWLKENAGAAS	GEFSKADEAGK	TIMESOLNKUM	WANRMDDIR	GIVQGAVNPFLMO	<b>.</b>
	orf29-1		!!!!!!!!!!! CEECDADEACE			CTIOCNER	ļ
45	01123 1	190	200				
		130	200	210 2	220	230 240	J
		250	260	270 2	280	290 300	n
	orf29a.pep	FQGVGIGAITDSAVSP				ODSAFAVKOGINS	9
		1111111111111111		1 1 11 1314			ı
50	orf29-1	FQGVGIGAITDSAVSP	VTDTAAQQTLQ	GINDLGKLSPE	AOLAAASLL	ODSAFAVKDGINS	Ś
		250				290 300	
		•	• .				
	•	310			340	350 360	0
	orf29a.pep	ARQWADAHPNITATAQ	TALAVAXAATT	VWGGKKVELNI	PTKWDWVKNT	GYXTPAVRTMHTI	Ĺ
55		1:11111111111111	111::1 11 1	11 1111111		11 11:1 1:11	ı
	orf29-1	AKQWADAHPNITATAQ				GYKKPAARHMQTI	Ĺ
		310	320	330 3	340	350 360	0
		224	200	200			_
60	0xf20= ===	370				410 420	0
JU	orf29a.pep	DGEMAGGNRPPKSITS		<b>LUAULIGEQIX</b>	CAGHAYNKHV	IKQQEFTDLNINS	S
	orf29-1	:   : DGEMAGGNKPIKSLP-	:	ekenicamaca a	ornomer —	DNADOTI COS	,
	Q1129-1	370	380		400	•	K
		310	300	330 .	100	410	

#### Homology with a predicted ORF from N.gonorrhoeae

ORF29 shows 88.8% identity over a 125aa overlap with a predicted ORF (ORF29.ng) from N. gonorrhoeae:

_	orf29.pep	VSPVLPITHERTGFEGVIGYETHFSGHGHE	30
3	orf29ng	: :	102
	orf29.pep	VHSPFDHHDSKSTSDFSGGVDGGFTVYQLHRTWSEIHPEDEYDGPQAAXYPPPGGARDIY	90
10	orf29ng	VHSPFDNHDSKSTSDFSGGVDGGFTVYQLHRTGSEIHPEDGYDGPQGGGYPPPGGARDIY	162
	orf29.pep	SYYVKGTSTKTKTSIVPQAPFSDRWLEENAGAASG	125
	orf29ng	SYHIKGTSTKTKINTVPQAPFSDRWLKENAGAASGFLSRADEAGKLIWENDPDKNWRANR	222

The complete length ORF29ng nucleotide sequence <SEQ ID 169> is predicted to encode a protein having amino acid sequence <SEQ ID 170>:

```
MNLPIQKFMM LFAAAISLLQ IPISHANGLD ARLRDDMQAK HYEPGGKYHL
                51
                    FGNARGSVKN RVCAVQTFDA TAVGPILPIT HERTGFEGVI GYETHFSGHG
               101
                    HEVHSPFDNH DSKSTSDFSG GVDGGFTVYQ LHRTGSEIHP EDGYDGPQGG
                    GYPPPGGARD IYSYHIKGTS TKTKINTVPQ APFSDRWLKE NAGAASGFLS
20
               151
                    RADEAGKLIW ENDPDKNWRA NRMDDIRGIV QGAVNPFLTG FQGLGVGAIT
               201
               251 DSAVSPVTYA AARKTLQGIH NLGNLSPEAQ LAAATALQDS AFAVKDSINS
                    ARQWADAHPN ITATAQTALA VTEAATTVWG GKKVELNPAK WDWVKNTGYK
               351 KPAARHMQTV DGEMAGGNKP LESKNTVTTN NFFENTGYTE KVLRQASNGD
25
               401
                    YHGFPQSVDA FSENGTVIQI VGGDNIVRHK LYIPGSYKGK DGNFEYIREA
               451 DGKINHRLFV PNQQLPEK*
```

In a second experiment, the following DNA sequence <SEQ ID 171> was identified:

	•	-+> > mmm.c.c	OM2.000.2.2.2.2			
	1	atgaarrige	CTATTCAAAA	ATTCATGATG	ctgttggcAg	cggcaatatc
20	51			GTCATGCGAA		
30	101			CACTACGAAC		
	151			TGTTAAAAAT		
	201			GCCCCATACT		
	251			GGCTATGAAA		
	301	CACGAAGTAC	ACAGTCCGTT	CGATAATCAT	GATTCAAAAA	GCACTTCTGA
35	351	TTTCAGCGGC	GGCGTAGACG	GCGGTTTTAC	CGTTTACCAA	CTTCATCGGA
	401	CAGGGTCGGA	AATACATCCC	GCAGACGGAT	ATGACGGGCC	TCAAGGCGGC
	451	GGTTATCCGG	AACCACAAGG	GGCAAGGGAT	ATATACAGCT	ACCATATCAA
	501	AGGAACTTCA	ACCAAAACAA	AGATAAACAC	TGTTCCGCAA	GCCCCTTTTT
	551	CAGACCGCTG	GCTAAAAGAA	AATGCCGGTG	CCGCTTCCGG	TTTTCTCAGC
40	601	CGTGCGGATG	AAGCAGGAAA	ACTGATATGG	GAAAACGACC	CCGATAAAAA
	651	TTGGCGGGCT	AACCGTATGG	ATGATATTCG	CGGCATCGTC	CAAGGTGCGG
	701	TTAATCCTTT	TTTAACGGGT	TTTCAAGGGG	TAGGGATTGG	GGCAATTACA
	751	GACAGTGCGG	TAAGCCCGGT	CACAGATACA	GCCGCTCAGC	AGACTCTACA
	801	AGGTATTAAT	GATTTAGGAA	ATTTAAGTCC	GGAAGCACAA	CTTGCCGCCG
45	851	CGAGCCTATT	ACAGGACAGT	GCCTTTGCGG	TAAAAGACGG	CATCAATTCC
	901	GCCAGACAAT	GGGCTGATGC	CCATCCGAAT	ATAACAGCAA	CAGCCCAAAC
	951	TGCCCTTGCC	GTAGCAGAGG	CCGCAGGTAC	GGTTTGGCGC	GGTAAAAAAG
	1001	TAGAACTTAA	CCCGACCAAA	TGGGATTGGG	TTAAAAATAC	CGGCTATAAA
	1051	AAACCTGCTG	CCCGCCATAT	GCAGACTGTA	GATGGGGAGA	TGGCAGGGGG
50	1101	GAATAGACCG	CCTAAATCTA	TAACGTCGGA	AGGAAAAGCT	AATGCTGCAA
	1151	CCTATCCTAA	GTTGGTTAAT	CAGCTAAATG	AGCAAAACTT	AAATAACATT
	1201	GCGGCTCAAG	ATCCAAGATT	GAGTCTAGCT	ATTCATGAGG	GTAAAAAAAA
	1251	TTTTCCAATA	GGAACTGCAA	CTTATGAAGA	GGCAGATAGA	CTAGGTAAAA
	1301			AGACAAACTA		
55	1351			TCGGCCACCA		
<del>-</del> -	1401			CAAATTTTGA		
	1451			AAAAATGGAC		

This encodes a protein having amino acid sequence <SEQ ID 172; ORF29ng-1>:

<sup>60 1</sup> MNLPIQKFMM LLAAAISMLH IPISHANGLD ARLRDDMQAK HYEPGGKYHL 51 FGNARGSVKN RVCAVQTFDA TAVGPILPIT HERTGFEGVI GYETHFSGHG

	101	HEVHSPFDNH	DSKSTSDFSG	GVDGGFTVYQ	LHRTGSEIHP	ADGYDGPQGG
	151	GYPEPQGARD	IYSYHIKGTS	TKTKINTVPQ	APFSDRWLKE	NAGAASGELS
	201	RADEAGKLIW	ENDPDKNWRA	NRMDDIRGIV	QGAVNPFLTG	FQGVGIGAIT
_	251	DSAVSPVTDT	AAQQTLQGIN	DLGNLSPEAQ	LAAASLLQDS	AFAVKDGINS
5	301	ARQWADAHPN	ITATAQTALA	VAEAAGTVWR	GKKVELNPTK	WDWVKNTGYK
	351	KPAARHMQTV	DGEMAGGNRP	PKSITSEGKA	NAATYPKLVN	QLNEQNLNNI
	401	AAQDPRLSLA	IHEGKKNFPI	GTATYEEADR	LGKIWVGEGA	ROTSGGGWLS
	451	RDGTRQYRPP	TEKKSQFATT	GIQANFETYT	IDSNEKRNKI	KNGHLNIR*

### ORF29ng-1 and ORF29-1 show 86.0% identity in 401 aa overlap:

10		10	20	30	. 40	50	60
	orf29ng-1.pep	MNLPIQKFMMLLAAA:	[SMLHIPISH	IANGLDARLR	DDMQAKHYEP	GGKYHLFGNA	RGSVKN
	orf29-1	MNLPIQKFMMLFAAA	ISLLQIPISH	ANGLDARLR	DDMQAKHYEP	GGKYHLFGNA'	IIIII: RGSVKK
15	•	10	20	30	40	50	60
	orf29ng-1.pep	70	80	90	100	110	120
	Olizang-i.pep	RVCAVQTFDATAVGP	:	111111111	1111111111	1111-11131	
20	orf29-1	RVYAVQTFDATAVSP	/LPITHERTO	FEGVIGYET	HFSGHGHEVH	SPFDHHDSKS	TSDFSG
20		70	80	90	100	110	120
	orf29ng-1.pep	130	140	150	160	170	180
	orrzang-r.pep	GVDGGFTVYQLHRTGS	PILL TILL	GPOGGGYPE	POGARDIYSY	HIKGTSTKTK	INTVPQ
25	orf29-1	GVDGGFTVYQLHRTG	SEIHPEDGYI	GPQGSDYPP	PGGARDIYSY	YVKGTSTKTK	TNIVPQ
		130	140	150	160	170	180
	* * * * * * * * * * * * * * * * * * * *	190	200	210	220	230	240
30	orf29ng-1.pep	APFSDRWLKENAGAAS	SGFLSRADE	GKLIWENDP	DKNWRANRMD	DIRGIVQGAV	NPFLTG
30	orf29-1	APFSDRWLKENAGAAS	GFFSRADE	GKLIWESDP	:          NKNWWANRMD	:          DVBGTVOGAVI	IIII I
		190	200	210	220	230	240
35		250	260	270	280	290	300
33	orf29ng-1.pep	FQGVGIGAITDSAVSI	PVTDTAAQQT	LQGINDLGN	LSPEAQLAAA	SLLQDSAFAVI	KDGINS
	orf29-1	FOGVGIGALTDSAVS	PVTDTAAQQT	LOGINDLGK	LSPEAQLAAA	SLLQDSAFAV	KDGINS
	•	250	260	270	280	290	300
40		310	320	330	340	350	360
	orf29ng-1.pep	ARQWADAHPNITATAC	TALAVAEAA	GTVWRGKKV	ELNPTKWDWV	KNTGYKKPAAI	RHMOTV
	orf29-1	i:                  AKQWADAHPNITATAG	: :         ממשממפ.זמינו	CTVMPCKKV			:
	, , , , , , , , , , , , , , , , , , ,	310	320	330	340	ANTGYKKPAAI 350	360
45		270					-
	orf29ng-1.pep	370 DGEMAGGNRPPKSI-1	380 SEGKANAAT	390 YPKLVNOLN	400 FONLNNIADO	410	419
	• • •		: :::::	: 1: :: :	:::::		
50	orf29-1	DGEMAGGNKPIKSLPN 370	ISAAEKRKQN 380	FEKFNSNWS			
30			300	390	400	410	420
		20 430	440	450	460	470	479
	orf29ng-1.pep	IGTATYEEADRLGKIV	VGEGARQTS	GGGWLSRDG	TRQYRPPTEK	KSQFATTGIQA	ANFETY
55	orf29-1	RYTSLDGKITIIKDNE	NNYFRIHDN	SRKQYLDSN	GNAVKTGNLO	GKOAKDYI.000	OTHIRN
		430	440	450	460	470	480

Based on this analysis, including the presence of a putative leader sequence in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 21

60

The following partial DNA sequence was identified in N.meningitidis <SEQ ID 173>:

```
1 ATGAAAAAC AAATCACCGC AGCCGTAATG ATGCTGTCTA TGATTGCCCC
    CGCAATGGCA AACGGCTTGG ACAATCAGGC ATTTGAAGAC CAAATGTTCC
51
```

101 ACACGCGGC AGATGCACCG ATGCAG...

This corresponds to the amino acid sequence <SEQ ID 174; ORF30>:

5 1 MKKQITAAVM MLSMIAPAMA NGLDNQAFED QMFHTRADAP MQ..

Further work revealed the complete nucleotide sequence <SEQ ID 175>:

```
1 ATGAAAAAC AAATCACCGC AGCCGTAATG ATGCTGTCTA TGATTGCCCC
                    CGCAATGGCA AACGGCTTGG ACAATCAGGC ATTTGAAGAC CAAGTGTTCC
                51
               101
                    ACACGCGGC AGATGCACCG ATGCAGTTGG CGGAGCTTTC TCAAAAGGAG
10
                    ATGAAGGAGA CAGAGGGGC GTTTCTTCCA TTGGCTATCT TGGGTGGTGC
               151
                    TGCCATTGGT ATGTGGACAC AGCATGGTTT TAGTTATGCA ACGACAGGCA
               201
                    GACCAGCTTC TGTTAGAGAT GTTGCTATTG CTGGCGGATT AGGCGCAATT
               251
                    CCTGGTGGTG TAGGCGCCGC AGGAAAGGTT GTTTCCTTTG CTAAATATGG
               301
                    ACGTGAGATT AAAATCGGCA ATAATATGCG GATAGCCCCT TTCGGTAATA
               351
15
                    GAACAGGTCA TCCTATTGGA AAATTTCCCC ATTATCATCG TCGAGTTACG
               401
                     GATAATACGG GCAAGACTTT GCCTGGACAG GGAATTGGTC GTCATCGCCC
                    TTGGGAATCA AAATCTACGG ACAGATCATG GAAAAACCGC TTCTAA
```

This corresponds to the amino acid sequence <SEQ ID 176; ORF30-1>:

```
MKKQITAAVM MLSMIAPAMA NGLDNQAFED QVFHTRADAP MQLAELSQKE
20
                     MKETEGAFLP LAILGGAAIG MWTQHGFSYA TTGRPASVRD VAIAGGLGAI
                51
               101
                     PGGVGAAGKV VSFAKYGREI KIGNNMRIAP FGNRTGHPIG KFPHYHRRVT
                     DNTGKTLPGQ GIGRHRPWES KSTDRSWKNR F*
```

Computer analysis of this amino acid sequence gave the following results:

### Homology with a predicted ORF from N. meningitidis (strain A)

25 ORF30 shows 97.6% identity over a 42aa overlap with an ORF (ORF30a) from strain A of N. meningitidis:

```
10
                                    20
                                            30
                    MKKQITAAVMMLSMIAPAMANGLDNQAFEDQMFHTRADAPMQ
         orf30.pep
                    30
         orf30a
                    MKKQITAAVMMLSMIAPAMANGLDNQAFEDQVFHTRADAPMQLAELSQKEMKXTXGAFLP
                                    20
                                            30
                                                     40
                                                              50
         orf30a
                    LXILGGAAIGMWTQHGFSYATTGRPASVRDVAIAGGLGAIPGXVGAAGKVVSFAKYGREI
                                                    100
```

35 The complete length ORF30a nucleotide sequence <SEQ ID 177> is:

```
ATGAAAAAAC AAATCACCGC AGCCGTAATG ATGCTGTCTA TGATTGCCCC
                51
                    CGCAATGGCA AACGGCTTGG ACAATCAGGC ATTTGAAGAC CAAGTGTTCC
                    ACACGCGGGC AGATGCACCG ATGCAGTTGG CGGAGCTTTC TCAAAAGGAG
               101
               151
                    ATGAAGGANA CAGNGGGGGC GTTTCTTCCA TTGGNTATCT TGGGTGGTGC
40
                    TGCCATTGGT ATGTGGACAC AGCATGGTTT TAGTTATGCA ACGACAGGCA
               201
                    GACCAGCTTC TGTTAGAGAT GTTGCTATTG CTGGCGGATT AGGCGCAATT
               251
                    CCTGGTGNTG TAGGCGCCGC AGGAAAGGTT GTTTCCTTTG CTAAATATGG
               301
               351
                    ACGTGAGATT AAAATCGGCA ATAATATGCG GATAGCCCCT TTCGGTAATA
                    GAACAGGTCA TCCTATTGGN AAATTTCCCC ATTATCATCG TCGAGTTACG
               401
45
               451
                    GATAATACGG GCAAGACTTT GCCTGGACAG GGAATTGGTC GTCATCGCCC
                   TTGGGAATCA AAATCTACGG ACAGATCATG GAAAAACCGC TTCTAA
               501
```

This encodes a protein having amino acid sequence <SEQ ID 178>:

```
MKKQITAAVM MLSMIAPAMA NGLDNQAFED QVFHTRADAP MQLAELSQKE
                    MKXTXGAFLP LXILGGAAIG MWTQHGFSYA TTGRPASVRD VAIAGGLGAI
                51
50
                    PGXVGAAGKV VSFAKYGREI KIGNNMRIAP FGNRTGHPIG KFPHYHRRVT
                    DNTGKTLPGQ GIGRHRPWES KSTDRSWKNR F*
```

ORF30a and ORF30-1 show 97.8% identity in 181 aa overlap:

orf30-1    MKKQITANVMLISHIAPANANGLUNQAFEQOYHTRADAPMQALSKYKKKTETAGEL 126   orf30a.pep		•		
orf30-1 LAILGGARGKWTOHESTATTGREANUNDAIAGGLGAIPGGWGANGKWVSFAKYGREI 120 orf30a.pep KIGNNMRIAPFGNRTGHPIGKPHYHRRVTDNTGKTLPGGGIGRHRPWESKSTDRSWKNR 180 orf30-1 EIGHMANIAPFGNRTGHPIGKPHYHRRVTDNTGKTLPGGGIGRHRPWESKSTDRSWKNR 180 orf30-1 FX  15  Homology with a predicted ORF from N.gonorrhoeae ORF30 shows 97.6% identity over a 42aa overlap with a predicted ORF (ORF30.ng) from N.gonorrhoeae:  20 orf30-1 PX  20 orf30-1 PX  10 orf30-1 PX  11 10 orf30ng MKKQITAAVMMLSHIAPRANANGLDNQAFEDQMPHYRADAPMQ orf30ng MKKQITAAVMSLSHIAPRANANGLDNQAFEDQWPHYRADAPMQ The complete length ORF30ng nucleotide sequence <seq 179="" id=""> is  1 ATGARAAAAA CAAATCACCGC AGCCGTAATG AGCGGCTTC TGAGAAGGGG 151 AGCGGCGAAG AACGGATTG ACAATCAGGC ATTGAGAGCCCC 251 GACCACGGGG AGCGGACGGG AGCGGTGGGGGACGTTC TGAGAAGGAG 151 AGCAGGGGGC AGAGGGGGGA AGCGACTGG AGCGGTATT TAGTGGGAGGG 151 AGCAGGGGG AGAGGGGGGA AGAGGGGGA AGCGACTGG AGCGGTATT TGAGGGGGGA 252 IGACCACCTTC TGTTSAGAGG ACAATCAGGC ATTGAGAGGA AGCGACTGGGGA AGCGACTGGGAAAGGGGA AGCGACTGGGGA AGCGACTGGGGGA AGCGACTGGGGA AGCGACTGGGGGA AGCGACTGGGGGGA AGCGACTGGGGGA AGCGACTGGGGGA AGCGACTGGGGGA AGCGACTGGGGGA AGCGACTGGGGGA AGCGACTGGGGGA AGCGGGACTGGGGGA AGCGACTGGGGGA AGCGACTGGGGGA AGCGACTGGGGGA AGCGACTGGGGGA</seq>		orf30-1		60
Orf30a.pep	5	orf30a.pep	LXILGGAAIGMWTQHGFSYATTGRPASVRDVAIAGGLGAIPGXVGAAGKVVSFAKYGREI	120
Orf30-1   KIGNNRIAPFONRTGPTGKFPHYRRYDNTGKTLPGGGIGRRPWESKSTDRSWKNR 180   Orf30-1   FX	,	orf30-1	LAILGGAAIGMWTQHGFSYATTGRPASVRDVAIAGGLGAIPGGVGAAGKVVSFAKYGREI	120
orf30a.pep FX orf30-1 FX  Homology with a predicted ORF from N.gonorrhoeae  ORF30 shows 97.6% identity over a 42aa overlap with a predicted ORF (ORF30.ng) from N.gonorrhoeae:  20 orf30-pep MKKQITAAVMMLSHIAPAMANGLDNQAFEDQWFHTRADAPMQ 42  orf30ng MKKQITAAVMMLSHIAPAMANGLDNQAFEDQWFHTRADAPMQ 42  The complete length ORF30ng nucleotide sequence <seq 179="" id=""> is  1 ATGAAAAAC AMATCACCGC ACCGTAATG ATGCTGTCTA TGATGGCCC 51 CGCAATGGCA AACGGATTG ACAATCAGGC ATTTGAACAC CAAGTGTTCC 51 CGCAATGGCA AACGGATTG ACAATCAGGC ATTTGACAC CAAGTGTTCC 201 TGCCATTGGT ATGTGACAC ACCAGTATT TAGTTATCCA ACGACAGGCA 215 ATCAAAGCAGA CTTTGCTCACAC AGCAGATTT TAGTTATCCA ACGACAGGCA 221 GACCAGCTT CTTTAGACAC ACCATGATT ATGTGATCAC ACGACAGGCA 221 GACCAGCTT CTTTAGACAC ACCATGATT ATGTGATCAC ACGACAGGCA 231 GACCAGCTT CTTTAGACAC ACCATGATT ATGTGATCAC ACGACAGGCA 231 GACCAGCTT CTTTAGACAC ACCATGATT ATGTGATCACAC 231 GACCAGCTT CTTTAGACACA ACCATGGTT TAGTTATCCA ACGACAGGCA 231 GACCAGCTT CTTTAGACACA ACCATGGTT TAGTTATCCA ACGACAGAC 231 GACCAGCTT CTTTAGACACA ACCATGGTT TAGTTATCACA ACGACAGAC 231 GACCAGCTT CTTTAGACACA ACCATGGTT TAGTTATCACA ACGACAGAC 231 GACCAGCTT CTTTAGACACA ACCATGGTT TAGTTATCACA ACCACAGAT ATGTGATCACAC 231 GACCAGCTT CTTTAGACACA ACCATGGTT TAGTTATCACACAC 231 GACCAGCTT CTTTAGACACA ACCATGGTT TAGTTATCACACACAC 231 GACCAGCTT CTTTAGACACA ACCATGGTT TAGTTATCACACACAC 231 GACCAGCTT CTTTAGACACACACACACACACACACACACACACACACAC</seq>		orf30a.pep	KIGNNMRIAPFGNRTGHPIGKFPHYHRRVTDNTGKTLPGQGIGRHRPWESKSTDRSWKNR	180
15	10	orf30-1	KIGNNMRIAPFGNRTGHPIGKFPHYHRRVTDNTGKTLPGQGIGRHRPWESKSTDRSWKNR	180
15		orf30a.pep	·	-
Homology with a predicted ORF from N.gonorrhoeae	15	orf30-1		•
20		Homology with a pre	dicted ORF from N.gonorrhoeae	
20		ORF30 shows 97.6%	identity over a 42aa overlap with a predicted ORF (ORF30.ng) fro	om N.
The complete length ORF30ng nucleotide sequence <seq 179="" id=""> is   The complete length ORF30ng nucleotide sequence <seq 179="" id=""> is   ATGAAAAAA AAATCACCGC ACCGTAATG ATGCTGTCTA TGATCGCCC CGCAATGGGCA AACGGATTGG ACAATCAGGC AATTGAAGAC CAAGTGTTCC ACAGCGGGC AAGGGATTGG ACAATCAGGC ATTGAAGAC CAAGTGTTCC ACAGCGGGC AGAGGGCTTTCTCCATTGGATCATC TGGGTGGTGC CTGAAGGAC CTAATGGAC CTAATGGGC TTTCTCCA TTGGATCATCT TGGGTGGTGC CTGACAAGGAC ACAGCGCTAT TGCAATCATC TGGTGGTGC CTGTTCAAGACGAC ACACCAGGCA AGACCAGGCA AGACCAGGCA AGACCAGGCA AGACCAGGCA AGACCAGGCA AGACCAGGCA AGACCAGGCA AGACCAGGCA CTAATCAGTT TGGAAAAATT TCCCATTATC ATGTTAGCA ACACCAGGCA AGACCAGGCA GAACAGCAGCA GAACAACCAG CTAATCATCT TGTAAAAATT TCCCAATAAT ATCGACAGGC AATCACCAGA GAACAACCA CTAACCATATC ATGTTCGGT AATCAAAAACA GAACAACCAA AACACCACTA ACACCAGGAA AGACCAGGCA AGACCAGCA AGACCAGCA AGACCAGGCA AGACCAGCA AG</seq></seq>				
The complete length ORF30ng nucleotide sequence <seq 179="" id=""> is    The complete length ORF30ng nucleotide sequence <seq 179="" id=""> is    The complete length ORF30ng nucleotide sequence <seq 179="" id=""> is    The complete length ORF30ng nucleotide sequence <seq 179="" id=""> is    The complete length ORF30ng nucleotide sequence <seq 179="" id=""> is    The complete length ORF30ng nucleotide sequence <seq 179="" id=""> is    The complete length ORF30ng nucleotide sequence <seq 179="" id=""> is    The complete length ORF30ng nucleotide sequence <seq 180="" id=""> is    The complete length ORF30ng nucleotide sequence <seq 180="" id=""> is    The complete length ORF30ng nucleotide sequence <seq 180="" id=""> is    The complete length ORF30ng nucleotide sequence <seq 180="" id=""> is    The complete length ORF30ng nucleotide sequence <seq 180="" id=""> is    This encodes a protein having amino acid sequence <seq 180="" id=""> is    This encodes a protein having amino acid sequence <seq 180="" id=""> is    The complete length ORF30ng nucleotide sequence <seq 180="" id=""> is    The complete length ORF30ng nucleotide sequence <seq 180="" id=""> is    The complete length ORF30ng nucleotide sequence <seq 180="" id=""> is    The complete length ORF30ng nucleotide sequence <seq 180="" id=""> is    The complete length ORF30ng nucleotide sequence <seq 180="" id=""> is    The complete length ORF30ng nucleotide sequence <seq 180="" id=""> is    The complete length ORF30ng nucleotide sequence <seq 180="" id=""> is    The complete length ORF30ng nucleotide sequence <seq 180="" id=""> is    The complete length of the complete sequence of the complete of the complete sequence of the complete of the complete sequence of the complet</seq></seq></seq></seq></seq></seq></seq></seq></seq></seq></seq></seq></seq></seq></seq></seq></seq></seq></seq></seq></seq></seq>	20		KKQITAAVMMLSMIAPAMANGLDNQAFEDQMFHTRADAPMQ	42
ATGAMARAC AMATCACCGC AGCCGTARTG ATGCTGTCTA TGATCGCCCC		orf30ng M	KKQITAAVMMLSMIAPAMANGLDNQAFEDQVFHTRADAPMQLAELSQKEMKETEGAFLP	60
10		The complete length	ORF30ng nucleotide sequence <seq 179="" id=""> is</seq>	
101		1 ATGA	AAAAAC AAATCACCGC AGCCGTAATG ATGCTGTCTA TGATCGCCCC	
1	25	101 ACAC	GCGGGC AGATGCGCCG ATGCAGTTGG CGGAGCTTTC TCAGAAGGAC	
301   GATGTAGGTG CTCCAGGAAA GGTTGTTTCC TTTGCTAAAT ATGGACGTGA   351   GATTAAAATC GCCATATAT ACGGACTAC CCCTTTGGT AATAGAACAG   401   GTCATCCTAT TGGAAAATT ACGCACTACT CATCGTCAGT TACGGATAAT   451   ACGGCAGAG CTTTGCCTGG ACAGGAATT GGTCGTCATC GCCCTTGGA   451   ACGGCAGAG CTTTGCCTGG ACAGGAATT GGTCGTCATC GCCCTTGGA   451   ACGGCACGAT CATGGAAAAA CCCCTTCTAA   451   ACGGCACGAT CATGGACAAAA CCCCTTCTAA   451   ACGGCACGAT CATGCACAAA CCCCTTCTAA   451   ACGGCACGAT CATGCACAAA CCCCTTCTAA   451   ACGGCACGAT CATGCACAAA CCCCTTCTAA   451   ACGGCACGAT CATGCACAAAA CCCCTTCTAA   451   ACGGCACGAT CATGCACAAAA CCCCTTCTAA   451   ACGGCACGATCAAAAA CCCCTTCTAA   451   ACGGCACGATCAAAAAAAAAAAAAAAAAAAAAAAAAAAA		201 TGCC	ATTGGT ATGTGGACAC AGCATGGTTT TAGTTATGCA ACGACAGGCA	
401	20	301 GATG	TAGGTG CTGCAGGAAA GGTTGTTTCC TTTGCTAAAT ATGGACGTGA	
### ACGGGCAAGA CTTTGCCTGG ACAGGGAATT GGTCGTCTC GCCCTTGGGA  ################################	30	401 GTCA	TCCTAT TGGAAAATTT CCCCATTATC ATCGTCGAGT TACGGATAAT	
This encodes a protein having amino acid sequence <seq 180="" id="">:  1</seq>		451 ACGG	GCAAGA CTTTGCCTGG ACAGGGAATT GGTCGTCATC GCCCTTGGGA	
MKETEGAFLP LAILGGAAIG MWTQHGFSYA TTGRPASVRD VAGGLGAIPG DVGAAGKVVS FAKYGREIKI GNNMRIAPFG NRTGHPIGKF PHYHRRVTDN TGKTLPGQGI GRHRPWESKS TDRSWKNRF*    ORF30ng and ORF30-1 show 98.3% identity in 181 aa overlap:    10		the second secon		
101 DVGAAGKVVS FAKYGREIKI GNNMRIAPFG NRTGHPIGKF PHYHRRVTDN 151 TGKTLPGQGI GRHRPWESKS TDRSWKNRF*  ORF30ng and ORF30-1 show 98.3% identity in 181 aa overlap:  10	35	1 MKKO	ITAAVM MLSMIAPAMA NGLDNQAFED QVFHTRADAP MQLAELSQKE	
ORF30ng and ORF30-1 show 98.3% identity in 181 aa overlap:  10 20 30 40 50 60 orf30ng.pep orf30-1		101 DVGA	AGKVVS FAKYGREIKI GNNMRIAPFG NRTGHPIGKF PHYHRRVTDN	
10			·	•
Orf30ng.pep	40	014 2012 014 21	10 00 00	
45    MKKQITAAVMMLSMIAPAMANGLDNQAFEDQVFHTRADAPMQLAELSQKEMKETEGAFLP	40	orf30ng.pep	MKKQITAAVMMLSMIAPAMANGLDNQAFEDOVFHTRADAPMOLAELSOKEMKETEGAFLP	
70 80 90 100 110  orf30ng.pep		orf30-1	MKKQITAAVMLSMIAPAMANGLDNQAFEDQVFHTRADAPMQLAELSQKEMKETEGAFLP	
Orf30ng.pep	45	•	20 10 10	
50 orf30-1 LAILGGAAIGMWTQHGFSYATTGRPASVRDVAIAGGLGAIPGGVGAAGKVVSFAKYGREI 70 80 90 100 110 120  120 130 140 150 160 170  KIGNNMRIAPFGNRTGHPIGKFPHYHRRVTDNTGKTLPGQGIGRHRPWESKSTDRSWKNR		orf30ng.pep	LAILGGAAIGMWTQHGFSYATTGRPASVRDVAGGLGAIPGDVGAAGKVVSFAKYGREI	
120   130   140   150   160   170	50	orf30-1	LAILGGAAIGMWTOHGFSYATTGRPASVRDVAIAGGLGAIPGGVGAAGKVVSFAKYGREI	
orf30ng.pep KIGNNMRIAPFGNRTGHPIGKFPHYHRRVTDNTGKTLPGQGIGRHRPWESKSTDRSWKNR			120	
130 140 150 160 170 180  180  orf30ng.pep FX			KIGNNMRIAPFGNRTGHPIGKFPHYHRRVTDNTGKTLPGOGIGRHRPWESKSTDRSWKNR	
180 orf30ng.pep FX 60	55	orf30-1	KIGNNMRIAPFGNRTGHPIGKFPHYHRRVTDNTGKTLPGQGIGRHRPWESKSTDRSWKNR	
orf30ng.pep FX		•	100 170 180	
	60		FX	
	00	orf30-1		

WO 99/24578

-150-

Based on this analysis, including the presence of a putative leader sequence in the gonococcal protein, it is predicted that the proteins from N.meningitidis and N.gonorrhoeae, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 22

10

The following partial DNA sequence was identified in N. meningitidis <SEO ID 181>: 5

```
ATGAATAAAA CTCTCTATCG TGTAATTTTC AACCGCAAAC GTGGGGCTGT
 1
 51
    GrTAGCCGTT GCTGAAACTA CCAAGCGCGA AGGTAAAAGC TGTGCCGATA
    GTGATTCAGG CAGCGCTCAT GTGAAATCTG TTCCTTTTGG TACTACTCAT
101
151
    GCACCTGTTT GTg.CGTTaC AAATATCTTT TCTTTTTCTT TATTGGGCTT
    TTCTTTATGT TTGGCTGTAG GtacGGyCAA TATTGCTTTT GCTGATGGCA
201
```

This corresponds to the amino acid sequence <SEQ ID 182; ORF31>:

- MNKTLYRVIF NRKRGAVXAV AETTKREGKS CADSDSGSAH VKSVPFGTTH
- 51 APVCXVTNIF SFSLLGFSLC LAVGTXNIAF ADGI..
- 15 Further work revealed a further partial nucleotide sequence <SEO ID 183>:

```
1 ATGAATAAAA CTCTCTATCG TGTAATTTTC AACCGCAAAC GTGGGGCTGT
                 51
                    GGTAGCCGTT GCTGAAACTA CCAAGCGCGA AGGTAAAAGC TGTGCCGATA
                101
                     GTGATTCAGG CAGCGCTCAT GTGAAATCTG TTCCTTTTGG TACTACTCAT
                     GCACCTGTTT GTCGTTCAAA TATCTTTTCT TTTTCTTTAT TGGGCTTTTC
                151
20
                201
                     TTTATGTTTG GCTGTAGGTA CGGCCAATAT TGCTTTTGCT GATGGCATT..
```

This corresponds to the amino acid sequence <SEQ ID 184; ORF31-1>:

- MNKTLYRVIF NRKRGAVVAV AETTKREGKS CADSDSGSAH VKSVPFGTTH
- APVCRSNIFS FSLLGFSLCL AVGTANIAFA DGI..

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from N.gonorrhoeae 25

> ORF31 shows 76.2% identity over a 84aa overlap with a predicted ORF (ORF31.ng) from N. gonorrhoeae:

```
orf31.pep
                   MNKTLYRVIFNRKRGAVXAVAETTKREGKSCADSDSGSAHVKSVPFGTTHAPVCXVTNIF
                                                                         60
                    30
                   MNKTLYRVIFNRKRGAVVAVAETTKREGKSCADSGSGSVYVKSVSFIPTH--
        orf31ng
                                                                         54
                    SFSLLGFSLCLAVGTXNIAFADGI
         orf31.pep
                                                                         R4
                    11 111111111111
         orf31ng
                    CFSALGFSLCLALGTVNIAFADGIITDKAAPKTQQATILQTGNGIPQVNIQTPTSAGVSV 114
```

35 The complete length ORF31ng nucleotide sequence <SEO ID 185> is:

```
ATGAACAAAA CCCTCTATCG TGTGATTTTC AACCGCAAAC GCGGTGCTGT
                51
                   GGTAGCTGTT GCCGAAACCA CCAAGCGCGA AGGTAAAAGC TGTGCCGATA
               101
                    GTGGTTCGGG CAGCGTTTAT GTGAAATCCG TTTCTTTCAT TCCTACTCAT
               151
                    TCCAAAGCCT TTTGTTTTTC TGCATTAGGC TTTTCTTTAT GTTTGGCTTT
40
                    GGGTACGGTC AATATTGCTT TTGCTGACGG CATTATTACT GATAAAGCTG
               201
               251 CTCCTAAAAC CCAACAAGCC ACGATTCTGC AAACAGGTaa cGGCATACCG
                    CAAGTCAATA TTCAAACCCC TACTTCGGCA GGGGTTTCTG TTAATCAATA
               301
               351
                    TGCCCAGTTT GATGTGGGTA ATCGCGGGGC GATTTTAAAC AACAGTCGCA
               401
                    GCAACACCCA AACACAGCTA GGCGGTTGGA TTCAAGGCAA TCCTTGGTTG
45
               451 ACAAGGGGCG AAGCACGTGT GGTTGTAAAC CAAATCAACA GCAGCCATCC
               501
                    TTCACAACTG AATGGCTATA TTGAAGTGGG TGGACGACGT GCAGAAGTCG
               551
                    TTATTGCCAA TCCGGCAGGG ATTGCAGTCA ATGGTGGTGG TTTTATCAAT
               601
                    GCTTCCCGTG CCACTTTGAC GACAGGCCAA CCGCAATATC AAGCAGGAGA
               651
                    CTTTAGCGGC TTTAAGATAA GGCAAGGCAA TGCTGTAATC GCCGGACACG
```

10

15

20

25

30

35

40

45

```
-151-
               GTTTGGATGC CCGTGATACC GATTTCACAC GTATTCTTGT ATGCCAACAA
          701
               AATCACCTTG ATCAGTACGG CCGAACAAGC AGGCATTCGT AA
This encodes a protein having amino acid sequence <SEQ ID 186>:
               MNKTLYRVIF NRKRGAVVAV AETTKREGKS CADSGSGSVY VKSVSFIPTH
               SKAFCFSALG FSLCLALGTV NIAFADGIIT DKAAPKTQQA TILQTGNGIP
           51
          101
               QVNIQTPTSA GVSVNQYAQF DVGNRGAILN NSRSNTQTQL GGWIQGNPWL
               TRGEARVVVN QINSSHPSQL NGYIEVGGRR AEVVIANPAG IAVNGGGFIN
          151
          201
               ASRATLTTGQ PQYQAGDFSG FKIRQGNAVI AGHGLDARDT DFTRILVCQQ
          251 NHLDQYGRTS RHS*
This gonococcal protein shares 50% identity over a 149aa overlap with the pore-forming
hemolysins-like HecA protein from Erwinia chrysanthemi (accession number L39897):
     orf31ng
              96 GNGIPQVNIQTPTSAGVSVNQYAQFDVGNRGAILNNSRSN-TQTQLGGWIQGNPWLTRGE 154
                  GNG+P VNI TP ++G+S N+Y F+V NRG ILNN + T +QLGG IQ NP L
                  GNGVPVVNIATPDASGLSHNRYHDFNVDNRGLILNNGTARLTPSQLGGLIQNNPNLNGRA 104
     HecA
     Orf31ng
              155 ARVVVNQINSSHPSQLNGYIEVGGRRAEVVIANPAGIAVNGGGFINASRATLTTGQPQYQ 214
                  A ++N++ S + S+L GY+EV G+ A VV+ANP GI +G GF+N R TLTTG PO+
     HecA
              105 AAAILNEVVSPNRSRLAGYLEVAGQAANVVVANPYGITCSGCGFLNTPRLTLTTGTPQFD 164
     Orf3lng 215 -AGDFSGFKIRQGNAVIAGHGLDARDTDF 242
                   AG SG +R G+ +I G GLDA +D+
     HecA
              165 AAGGLSGLDVRGGDILIDGAGLDASRSDY 193
Furthermore, ORF31ng and ORF31-1 show 79.5% identity in 83 aa overlap:
                                    20
                  MNKTLYRVIFNRKRGAVVAVAETTKREGKSCADSDSGSAHVKSVPFGTTHAPVCRSNIFS
     orf31-1.pep
                   MNKTLYRVIFNRKRGAVVAVAETTKREGKSCADSGSGSVYVKSVSFIPTH-
     orf31ng
                                                                        --SKAFC
                                    20
                                              30
                           70
                   FSLLGFSLCLAVGTANIAFADGI
     orf31-1.pep
                   11 [1111111:11:11111111
                   FSALGFSLCLALGTVNIAFADGIITDKAAPKTQQATILQTGNGIPQVNIQTPTSAGVSVN
     orf31ng
                                          80
                                                             100
On this basis, including the homology with hemolysins, and also with adhesins, it is predicted that
the proteins from N. meningitidis and N. gonorrhoeae, and their epitopes, could be useful antigens
for vaccines or diagnostics, or for raising antibodies.
Example 23
 The following partial DNA sequence was identified in N. meningitidis <SEQ ID 187>:
               ATGAATACTC CTCCTTTTGT CTGTTGGATT TTTTGCAAGG TCATCGACAA
            51
               TTTCGGCGAC ATCGGCGTTT CGTGGCGGCT CGCCCGTGTT TTGCACCGCG
           101
               AACTCGGTTG GCAGGTGCAT TTGTGGACGG ACGATGTGTC CGCCTTGCGT
               GCGCTTTGCC CTGATTTGCC CGATGTTCCC TGCGTTCATC AGGATATTCA
           151
               TGTCCGCACT TGGCATTCCG ATGCGGCAGA TATTGATACC GCG..
 This corresponds to the amino acid sequence <SEQ ID 188; ORF32>:
               MNTPPFVCWI FCKVIDNFGD IGVSWRLARV LHRELGWQVH LWTDDVSALR
            51 ALCPDLPDVP CVHQDIHVRT WHSDAADIDT A..
```

ATGAATACTC CTCCTTTTGT CTGTTGGATT TTTTGCAAGG TCATCGACAA 50 TTTCGGCGAC ATCGGCGTTT CGTGGCGGCT CGCCCGTGTT TTGCACCGCG 101 AACTCGGTTG GCAGGTGCAT TTGTGGACGG ACGATGTGTC CGCCTTGCGT

Further work revealed the complete nucleotide sequence <SEQ ID 189>:

701

751

60

```
GCGCTTTGCC CTGATTTGCC CGATGTTCCC TGCGTTCATC AGGATATTCA
                    TGTCCGCACT TGGCATTCCG ATGCGGCAGA TATTGATACC GCGCCTGTTC
               201
                    CCGATGTCGT CATCGAAACT TTTGCCTGCG ACCTGCCCGA AAATGTGCTG
               251
                    CACATTATCC GCCGACACAA GCCGCTTTGG CTGAATTGGG AATATTTGAG
               301
                    CGCGGAGGAA AGCAATGAAA GGCTGCATCT GATGCCTTCG CCGCAGGAGG
5
               351
                     GTGTTCAAAA ATATTTTTGG TTTATGGGTT TCAGCGAAAA AAGCGGCGGG
               401
                     TTGATACGCG AACGTGATTA CTGCGAAGCC GTCCGTTTCG ATACTGAAGC
                451
                     CCTGCGAGAG CGGCTGATGC TGCCCGAAAA AAACGCCTCC GAATGGCTGC
               501
                     TTTTCGGCTA TCGGAGCGAT GTTTGGGCAA AGTGGCTGGA AATGTGGCGA
               551
                     CAGGCAGGCA GCCCGATGAC ACTGTTGCTG GCGGGGACGC AAATCATCGA
10
                601
                     CAGCCTCAAA CAAAGCGGCG TTATTCCGCA AGATGCCCTG CAAAACGACG
                651
                     GCGATGTTTT TCAGACGGCA TCCGTCCGCC TCGTCAAAAT CCCTTTCGTG
                701
                     CCGCAACAGG ACTTCGACCA ACTGCTGCAC CTTGCCGACT GCGCCGTCAT
                751
                     CCGCGGCGAA GACAGTTTCG TGCGCGCCCA GCTTGCGGGC AAACCCTTCT
                801
                     TTTGGCACAT CTACCCGCAA GACGAGAATG TCCATCTCGA CAAACTCCAC
15
                851
                     GCCTTTTGGG ATAAGGCACA CGGTTTCTAC ACGCCCGAAA CCGTGTCGGC
                901
                     ACACCGCCGT CTTTCGGACG ACCTCAACGG CGGAGAGGCT TTATCCGCAA
                951
                     CACAACGCCT CGAATGTTGG CAAACCCTGC AACAACATCA AAACGGCTGG
               1001
                     CGGCAAGGCG CGGAGGATTG GAGCCGTTAT CTTTTCGGGC AGCCGTCAGC
               1051
                     TCCTGAAAAA CTCGCTGCCT TTGTTTCAAA GCATCAAAAA ATACGCTAG
20
     This corresponds to the amino acid sequence <SEQ ID 190; ORF32-1>:
                  1 MNTPPFVCWI FCKVIDNFGD IGVSWRLARV LHRELGWQVH LWTDDVSALR
                     ALCPDLPDVP CVHQDIHVRT WHSDAADIDT APVPDVVIET FACDLPENVL
                     HIIRRHKPLW LNWEYLSAEE SNERLHLMPS PQEGVQKYFW FMGFSEKSGG
                101
                     LIRERDYCEA VRFDTEALRE RLMLPEKNAS EWLLFGYRSD VWAKWLEMWR
25
                151
                     QAGSPMTLLL AGTQIIDSLK QSGVIPQDAL QNDGDVFQTA SVRLVKIPFV
                201
                     POODFDOLLH LADCAVIRGE DSFVRAQLAG KPFFWHIYPO DENVHLDKLH
                251
                     AFWDKAHGFY TPETVSAHRR LSDDLNGGEA LSATQRLECW QTLQQHQNGW
                301
                     RQGAEDWSRY LFGQPSAPEK LAAFVSKHQK IR*w
      Computer analysis of this amino acid sequence gave the following results:
30
      Homology with a predicted ORF from N. meningitidis (strain A)
      ORF32 shows 93.8% identity over a 81aa overlap with an ORF (ORF32a) from strain A of N.
      meningitidis:
                                                                         50
                                                               40
                                                     30
                         MNTPPFVCWIFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDDVSALRALCPDLPDVP
 35
            orf32.pep
                                   MNTPPFSAGXFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDDVSALRALCPDLPDVX
            orf32a
                                                                         50
                                                               40
                                           20
                                                     30
                                           80
 40
                         CVHQDIHVRTWHSDAADIDTA
            orf32.pep
                         11111111111111111111111
                         CVHQDIHVRTWHSDAADIDTAPVXDVVIETFACDLPENVLHIIRRHKPLWLXWEYLSAEX
            orf32a
                                                              100
                                                                        110
                                           80
                                                     90
                                 70
       The complete length ORF32a nucleotide sequence <SEQ ID 191> is:
 45
                      ATGAATACTC CTCCTTTTC TGCTGGANTT TTTTGCAAGG TCATCGACAA
                      TTTCGGCGAC ATCGGCGTTT
                                            CGTGGCGGCT TGCCCGTGTT TTGCACCGCG
                  51
                      AACTCGGTTG GCAGGTGCAT TTGTGGACGG ACGATGTGTC CGCCTTGCGT
                 101
                      GCGCTTTGCC CTGATTTGCC CGATGTTCNC TGCGTTCATC AGGATATTCA
                 151
                      TGTCCGCACT TGGCATTCCG ATGCGGCAGA TATTGATACC GCGCCTGTTC
  50
                 201
                      NCGATGTCGT CATCGAAACT TTTGCCTGCG ACCTGCCCGA AAATGTGCTG
                 251
                      CACATCATCC GCCGACACAA GCCGCTTTGG CTGAANTGGG AATATTTGAG
                 301
                      CGCGGAGGAN AGCAATGAAA GGCTGCACNT GATGCCTTCG CCGCAGGAGA
                 351
                      GTGTTCNAAA ATANTTTTGG TTTATGGGTT TCAGCGAANN NAGCGGCGGA
                  401
                      CTGATACGCG AACGCGATTA CTGCGAAGCC GTCCGTTTCG ATAGCGGAGC
  55
                  451
                      CTTGCGCAAG AGGCTGATGC TTCCCGAAAA AAACGNCCCC GAATGGCTGC
                  501
                      TTTTCGGCTA TCGGAGCGAT GTTTGGGCAA AGTGGCTGGA AATGTGGCGA
                  551
                      CAGGCAGGCA GTCCGTTGAC ACTTTTGCTG GCNGGGGCGC ANATTATCGA
                  601
                      CAGCCTCAAA CAAAACGGCG TTATTCCGCA AGATGCCCTG CAAAACGACG
```

GCGATGTTTT TCAGACGGCA TCCGTCCGCC TCGTCAAAAT CCCTTTCGTG

CCGCAACAGG ACTTCGACAA ACTGCTGCAC CTTGCCGACT GCGCCGTCAT

```
CCGCGGCGAA GACAGTTTCG TGCGCGCCCA GCTTGCGGGC AAACCCTTCT
                  TTTGGCACAT CTACCCGCAA GATGAGAATG TCCATCTCGA CAAACTCCAC
              851
              901
                  GCCTTTTGGG ATAAGGCACA CGGTTTCTAC ACGCCCGAAA CCGCATCGGC
                  ACACCGCCGC CTTTCAGACG ACCTCAACGG CGGAGAGGCT TTATCCGCAA
              951
5
             1001
                  CACAACGCCT CGAATGTTGG CAAATCCTGC AACAACATCA AAACGGCTGG
             1051
                  CGGCAAGGCG CGGAGGATTG GAGCCGTTAT CTTTTTGGGC AGCCTTCCGC
             1101
                  ATCCGAAAAA CTCGCCGCCT TTGTTTCAAA GCATCAAAAA ATACGCTAG
    This encodes a protein having amino acid sequence <SEQ ID 192>:
                  MNTPPFSAGX FCKVIDNFGD IGVSWRLARV LHRELGWQVH LWTDDVSALR
10
               51
                  ALCPDLPDVX CVHQDIHVRT WHSDAADIDT APVXDVVIET FACDLPENVL
                  HIIRRHKPLW LXWEYLSAEX SNERLHXMPS PQESVXKXFW FMGFSEXSGG
              101
                  LIRERDYCEA VRFDSGALRK RLMLPEKNXP EWLLFGYRSD VWAKWLEMWR
              151
              201
                  QAGSPLTLLL AGAXIIDSLK QNGVIPQDAL QNDGDVFQTA SVRLVKIPFV
              251
                  POODFDKLLH LADCAVIRGE DSFVRAQLAG KPFFWHIYPO DENVHLDKLH
15
              301
                  AFWDKAHGFY TPETASAHRR LSDDLNGGEA LSATQRLECW QILQQHQNGW
              351
                  RQGAEDWSRY LFGQPSASEK LAAFVSKHQK IR*
    ORF32a and ORF32-1 show 93.2% identity in 382 aa overlap:
                                     20
                                              30
                     MNTPPFVCWIFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDDVSALRALCPDLPDVP
         orf32-1.pep
20
                     \Pi\Pi\Pi\Pi
                              orf32a
                     MNTPPFSAGXFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDDVSALRALCPDLPDVX
                                     20
                                              30
                                                       40
                                                               50
                            70
                                     80
                                              90
                                                      100
                                                              110
                                                                       120
25
                     CVHQDIHVRTWHSDAADIDTAPVPDVVIETFACDLPENVLHIIRRHKPLWLNWEYLSAEE
         orf32-1.pep
                     orf32a
                     CVHQDIHVRTWHSDAADIDTAPVXDVVIETFACDLPENVLHIIRRHKPLWLXWEYLSAEX
                            70
                                     80
                                              90
                                                      100
                                                              110
                                                                       120
30
                           130
                                    140
                                             150
                                                      160
                                                              170
                     SNERLHLMPSPQEGVQKYFWFMGFSEKSGGLIRERDYCEAVRFDTEALRERLMLPEKNAS
         orf32-1.pep
                     SNERLHXMPSPQESVXKXFWFMGFSEXSGGLIRERDYCEAVRFDSGALRKRLMLPEKNXP
          orf32a
                           130
                                    140
                                             150
                                                      160
                                                              170
                                                                       180
35
                           190
                                    200
                                             210
                                                      220
                     EWLLFGYRSDVWAKWLEMWRQAGSPMTLLLAGTQIIDSLKQSGVIPQDALQNDGDVFQTA
                     orf32a
                     EWLLFGYRSDVWAKWLEMWRQAGSPLTLLLAGAXIIDSLKQNGVIPQDALQNDGDVFQTA
40
                           190
                                    200
                                             210
                                                      220
                                                              230
                           250
                                    260
                                             270
                                                      280
                                                               290
          orf32-1.pep
                     SVRLVKIPFVPQQDFDQLLHLADCAVIRGEDSFVRAQLAGKPFFWHIYPODENVHLDKLH
                     45
          orf32a
                     SVRLVKIPFVPQQDFDKLLHLADCAVIRGEDSFVRAQLAGKPFFWHIYPQDENVHLDKLH
                           250
                                    260
                                             270
                                                      280
                                                              290
                                                                       300
                                    320
                                             330
                                                      340
                                                               350
                     {\tt AFWDKAHGFYTPETVSAHRRLSDDLNGGEALSATQRLECWQTLQQHQNGWRQGAEDWSRY}
          orf32-1.pep
 50
                     AFWDKAHGFYTPETASAHRRLSDDLNGGEALSATQRLECWQILQQHQNGWRQGAEDWSRY
          orf32a
                           310
                                    320
                                             330
                                                      340
                           370
                                    380
 55
                     LFGQPSAPEKLAAFVSKHQKIRX
          orf32-1.pep
                     LFGQPSASEKLAAFVSKHQKIRX
          orf32a
                           370
```

Homology with a predicted ORF from N.gonorrhoeae 

> ORF32 shows 95.1% identity over a 82aa overlap with a predicted ORF (ORF32.ng) from N. gonorrhoeae:

MNTPPF-VCWIFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDDVSALRALCPDLP orf32.pep 

WO 99/24578 PCT/IB98/01665

-154-

```
MVMNTYAFPVCWIFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDDVSALRALCPDLP
          orf32ng
                        DVPCVHQDIHVRTWHSDAADIDTA
                                                                                       81
          orf32.pep
                        111 111111111111111111111
 5
                        DVPFVHQDIHVRTWHSDAADIDTAPVPDAVIETFACDLPENVLNIIRRHKPLWLNWEYLS 120
          orf32ng
     An ORF32ng nucleotide sequence <SEQ ID 193> was predicted to encode a protein having amino
     acid sequence <SEQ ID 194>:
                     MVMNTYAFPV CWIFCKVIDN FGDIGVSWRL ARVLHRELGW QVHLWTDDVS
                 1
                     ALRALCPDLP DVPFVHQDIH VRTWHSDAAD IDTAPVPDAV IETFACDLPE
10
                     NVLNIIRRHK PLWLNWEYLS AEESNERLHL MPSPQEGVQK YFWFMGFSEK
               101
                     SGGLIRERDY REAVREDTEA LRRRLVLPEK NAPEWLLFGY RGDVWAKWLD
               201
                     MWQQAGSLMT LLLAGAQIID SLKQSGVIPQ NALQNEGGVF QTASVRLVKI
                251
                     PFVPQQDFDK LLHLADCAVI RGEDSFVRTQ LAGKPFFWHI YPQDENVHLD
                     KLHAFWDKAY GFYTPETASV HRLLSDDLNG GEALSATORL ECGVL*
               301
15
     Further sequencing revealed the following DNA sequence <SEQ ID 195>:
                 1 ATGAATACAT ACGCTTTTCC TGTCTGTTGG ATTTTTTGCA AGGTCATCGA
CAATTTCGGC GACATCGGCG TTTCGTGGCG GCTCGCCCGT GTTTTGCACC
                     GCGAACTCGG TTGGCAGGTG CATTTGTGGA CGGACGACGT GTCCGCCTTG
                101
                151
                     CGCGCGCTTT GTCCCGATTT GCCCGATGTT CCCTTCGTTC ATCAGGATAT
20
                201
                     TCATGTCCGC ACTTGGCATT CCGATGCGGC AGACATTGAT ACCGCGCCCG
                251
                     TTCCCGATGC CGTTATCGAA ACTTTTGCCT GCGACCTGCC CGAAAATGTG
                     CTGAACATCA TCCGCCGACA CAAACCGCTT TGGCTGAATT GGGAATATTT
                301
                351
                     GAGCGCGGAG GAAAGCAATG AAAGGCTGCA CCTGATGCCT TCGCCGCAGG
                401
                     AGGGCGTTCA AAAATATTTT TGGTTTATGG GTTTCAGCGA AAAAAGCGGC
25
                451
                     GGGTTGATAC GCGAACGCGA TTACCGCGAA GCCGTCCGTT TCGATACCGA
                     AGCCCTGCGC CGGCGGCTGG TGCTGCCCGA AAAAAACGCC CCCGAATGGC
                501
                551
                     TGCTTTTCGG CTATCGGGGC GATGTTTGGG CAAAGTGGCT GGACATGTGG
                601
                     CAACAGGCAG GCAGCCTGAT GACCCTACTG CTGGCGGGGG CGCAAATTAT
                     CGACAGCCTC AAACAAAGCG GCGTTATTCC GCAAAACGCC CTGCAAAAtg
                651
30
                     aaggcgGTGT CTTTCagacG gcatccgTcC gccttGTCAA AAtcCCGTTC
                701
                751
                     GTGCcGCAAC AGGACTTCGA CAAATTGCTG CAcctcgcCG ACTGCGCCGT
                801
                     GATACGCGGC GAAGACAGTT TCGTGCGTAC CCAGCTTGCC GGAAAACCCT
                851
                     TTTTTTGGCA CATCTACCCG CAAGACGAGA ATGTCCATCT CGACAAACTC
                     CACGCCTTTT GGGATAAGGC ATACGGCTTC TACACGCCCG AAACCGCATC
                901
35
                951
                     GGTGCACCGC CTCCTTTCGG ACGACCTCAA CGGCGGAGAG GCTTTATCCG
               1001
                     CAACACAACG CCTCGAATGT TGGCAAACCC TGCAACAACA TCAAAACGGC
               1051
                     TGGCGGCAAG GCGCGGAGGA TTGGAGCCGT TATCTTTTCG GGCAGCCTTC
                     CGCATCCGAA AAACTCGCCG CCTTTGTTTC AAAGCATCAA AAAATACGCT
               1101
               1151
      This encodes a protein having amino acid sequence <SEQ ID 196; ORF32ng-1>:
40
                     MNTYAFPVCW IFCKVIDNFG DIGVSWRLAR VLHRELGWQV HLWTDDVSAL
                 51
                     RALCPDLPDV PFVHQDIHVR TWHSDAADID TAPVPDAVIE TFACDLPENV
                101
                     LNIIRRHKPL WLNWEYLSAE ESNERLHLMP SPQEGVQKYF WFMGFSEKSG
                151
                     GLIRERDYRE AVRFDTEALR RRLVLPEKNA PEWLLFGYRG DVWAKWLDMW
45
                201
                     QQAGSLMTLL LAGAQIIDSL KQSGVIPQNA LQNEGGVFQT ASVRLVKIPF
                251
                     VPQQDFDKLL HLADCAVIRG EDSFVRTQLA GKPFFWHIYP QDENVHLDKL
                301
                     HAFWDKAYGF YTPETASVHR LLSDDLNGGE ALSATQRLEC WQTLQQHQNG
                     WRQGAEDWSR YLFGQPSASE KLAAFVSKHQ KIR*
      ORF32ng-1 and ORF32-1 show 93.5% identity in 383 aa overlap:
50
                                                       30
                          MNTPPF-VCWIFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDDVSALRALCPDLPDV
           orf32-1.pep
                              MNTYAFPVCWIFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDDVSALRALCPDLPDV
           orf32ng-1
                                  10
                                            20
                                                      30
                                                                          50
 55
                                             80
                                                       90
                                                                100
                                                                          110
                          PCVHQDIHVRTWHSDAADIDTAPVPDVVIETFACDLPENVLHIIRRHKPLWLNWEYLSAE
           orf32-1.pep
                          orf32ng-1
                          PFVHQDIHVRTWHSDAADIDTAPVPDAVIETFACDLPENVLNIIRRHKPLWLNWEYLSAE
 60
                                            80
                                  70
                                                      90
                                                               100
                                                                         110
                                                                                   120
                        120
                                  130
                                            140
                                                      150
                                                                160
                                                                          170
                                                                                   179
```

•	orf32-1.pep	ESNER	LHLMPSPQEG	VOKYFWFMGE	SEKSGGLIR	ERDYCEAVRF	DTEALRERLMI	LPEKNA
	orf32ng-1	ESNER	LHLMPSPQEG 130	VQKYFWFMGE	SEKSGGLIR 150	 ERDYREAVRF  160		
5	٠		•				170	180
	orf32-1.pep	180 SEWLL	190 FGYRSDVWAK	200 WLEMWRQAGS	210 PMTLLLAGT	QIIDSLKQSG <sup>1</sup>	230 VIPODALONDO	239 DVFOT
	orf32ng-1		1114:11141	:  :		QIIDSLKQSG	1111:1111:1	HH
10			190	200	210	220	230	240
	620 1	240	250	260	270	280	290	299
	orf32-1.pep	ASVRL	VKIPFVPQQD: 	FDQLLHLADO	AVIRGEDSF	VRAQLAGKPF	FWHIYPQDENV	HLDKL
15	orf32ng-1	ASVRL	VKIPFVPQQD: 250	FDKLLHLADO 260	AVIRGEDSF 270	VRTOLAGKPF1 280	FWHIYPODENV 290	HLDKL 300
		300	310	320	330	340	350	359
20	orf32-1.pep		Kahgfytpet'   :		LNGGEALSA'	TQRLECWQTL(	QQHQNGWRQGA	AEDWSR
	orf32ng-1	HAFWD	KAYGFYTPET 310	ASVHRLLSDI 320	LNGGEALSA 330	TORLECWOTL(		EDWSR
		•.				340	350	360
25	orf32-1.pep	360 YLFGQ	370 PSAPEKLAAF !!!!!!!!	380 VSKHQKIRX				
•	orf32ng-1		PSASEKLAAF 370				!	

On this basis, including the RGD sequence in the gonococcal protein, characteristic of adhesins, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF32-1 (42kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 7A shows the results of affinity purification of the His-fusion protein, and Figure 7B shows the results of expression of the GST-fusion in *E.coli*. Purified His-fusion protein was used to immunise mice, whose sera were used for ELISA, giving a positive result. These experiments confirm that ORF32-1 is a surface-exposed protein, and that it is a useful immunogen.

#### Example 24

35

40 The following partial DNA sequence was identified in N. meningitidis <SEQ ID 197>:

```
..TTGTTCCTGC GTGTNAAAGT GGGGCGTTTT TTCAGCAGTC CGGCGACGTG
                51
                      GTTTCGGGNC AAAGACCCTG TAAATCAGGC GGTGTTGCGG CTGTATNCGG
                      ACGAGTGGCG GCA.ACTTCG GTACGTTGGA AAATAGNCGC AACGTCGCAC
               101
               151
                      AGCCTGTGGC TCTGCACGCT GCTCGGAATG CTGGTGTCGG TATTGTTGCT
45
                      GCTTTTGGTG CGGCAATATA CGTTCAACTG GGAAAGCACG CTGTTGAGCA
               201
                      ATGCCGCTTC GGTACGCGCG GTGGAAATGT TGGCATGGCT GCCGTCGAAA
               251
                      CTCGGTTTCC CTGTCCCCGA TGCGCGGTCG GTCATCGAAG GCCGTCTGAA
               301
               351
                      CGGCAATATT GCCGATGCGC GGGCTTGGTC GGGGCTGCTG GTCGNCAGTA
                      TCGCCTGCTA NGGCATCCTG CCGCGCCTG..
```

50 This corresponds to the amino acid sequence <SEQ ID 198; ORF33>:

- 1 ..LFLRVKVGRF FSSPATWFRX KDPVNQAVLR LYXDEWRXTS VRWKIXATSH 51 SLWLCTLLGM LVSVLLLLLV RQYTFNWEST LLSNAASVRA VEMLAWLPSK
- 101 LGFPVPDARS VIEGRLNGNI ADARAWSGLL VXSIACXGIL PRL..

### Further work revealed the complete nucleotide sequence <SEQ ID 199>:

	1	ATGTTGAATC	CATCCCGAAA	ACTGGTTGAG	CTGGTCCGTA	TTTTGGACGA
	51	AGGCGGTTTT	ATTTTCAGCG	GCGATCCCGT	ACAGGCGACG	GAGGCTTTGC
	101	GCCGCGTGGA	CGGCAGTACG	GAGGAAAAA	TCATCCGTCG	GGCGGAGATG
5	151	ATTGACAGGA	ACCGTATGCT	GCGGGAGACG	TTGGAACGTG	TGCGTGCGGG
	201	GTCGTTCTGG	TTGTGGGTGG	TGGCGGCGAC	GTTTGCATTT	TTTACCGGTT
	251	TTTCAGTCAC	TTATCTTCTA	ATGGACAATC	AGGGTCTGAA	TTTCTTTTTG
	301	GTTTTGGCGG	GCGTGTTGGG	CATGAATACG	CTGATGCTGG	CAGTATGGTT
	351	GGCAATGTTG	TTCCTGCGTG	TGAAAGTGGG	GCGTTTTTTC	AGCAGTCCGG
10	401	CGACGTGGTT	TCGGGGCAAA	GACCCTGTAA	ATCAGGCGGT	GTTGCGGCTG
	451	TATGCGGACG	AGTGGCGGCA	ACCTTCGGTA	CGTTGGAAAA	TAGGCGCAAC
	501	GTCGCACAGC	CTGTGGCTCT	GCACGCTGCT	CGGAATGCTG	GTGTCGGTAT
	551	TGTTGCTGCT	TTTGGTGCGG	CAATATACGT	TCAACTGGGA	AAGCACGCTG
	601			ACGCGCGGTG		
15	651	GTCGAAACTC	GGTTTCCCTG	TCCCCGATGC	GCGGGCGGTC	ATCGAAGGCC
	701	GTCTGAACGG	CAATATTGCC	GATGCGCGGG	CTTGGTCGGG	GCTGCTGGTC
	751	GGCAGTATCG	CCTGCTACGG	CATCCTGCCG	CGCCTGCTGG	CTTGGGTAGT
	801	GTGTAAAATC	CTTTTGAAAA	CAAGCGAAAA	CGGATTGGAT	TTGGAAAAGC
	851	CCTATTATCA	GGCGGTCATC	CGCCGCTGGC	AGAACAAAAT	CACCGATGCG
20	901	GATACGCGTC	GGGAAACCGT	GTCCGCCGTT	TCACCGAAAA	TCATCTTGAA
	951			TCATGCTGGA		
	1001			GCGCAGGAAT		
	1051			CGCGCTGGAG		
	1101			TGCGCGCCCA		
25	1151			CTCTCGGAAG		
	1201			GGGGCTTTCA		
	1251			TGGCCGAATG		
	1301	CTGACAGGGC	GGCGCAGGAA	GGGCGTTTGA	<b>AAGACCAATA</b>	A

#### This corresponds to the amino acid sequence <SEQ ID 200; ORF33-1>:

30	1	MLNPSRKLVE	LVRILDEGGF	IFSGDPVQAT	EALRRVDGST	EEKIIRRAEM
	51	IDRNRMLRET	LERVRAGSFW	LWVVAATFAF	FTGFSVTYLL	MDNQGLNFFL
	101	VLAGVLGMNT	LMLAVWLAML	FLRVKVGRFF	SSPATWFRGK	DPVNQAVLRL
	151			LWLCTLLGML		
	201	LSNAASVRAV	EMLAWLPSKL	GFPVPDARAV	IEGRLNGNIA	DARAWSGLLV
35	251	GSIACYGILP	RLLAWVVCKI	LLKTSENGLD	LEKPYYQAVI	RRWQNKITDA
	301	DTRRETVSAV	SPKIILNDAP	KWAVMLETEW	QDGEWFEGRL	AQEWLDKGVA
	351	TNREQVAALE	TELKQKPAQL	LIGVRAQTVP	DRGVLRQIVR	LSEAAQGGAV
	401	VQLLAEQGLS	DDLSEKLEHW	RNALAECGAA	WLEPDRAAQE	GRLKDQ*

Computer analysis of this amino acid sequence gave the following results:

### 40 Homology with a predicted ORF from N.meningitidis (strain A)

ORF33 shows 90.9% identity over a 143aa overlap with an ORF (ORF33a) from strain A of N. meningitidis:

45	orf33.pep				LFLRV	10 KVGRFFSSPA	20 TWFRXKDPVI	30 NQAVLR
43	orf33a	T MDNOC	NEET UT ACT	IV CMMINIT MIT BI	TAL DATE DI DA	MACD DECODE		111111
	OLLDDA			/XGMNTLMLAV				NOAVLR
		90	100	110	120	130	140	
			40	50	60	70	80	90
50	orf33.pep	LYXDEW	RXTSVRWKIX	KATSHSLWLCT	TLLGMLVSVL	LLLLVROYTE	NWESTLLSNA	AASVRA
	• •	11 111	11 111111	1111111111			1111111::	
	orf33a	LYADEW	RXPSVRWKIC	ATSHSLWLC	LLGMLVSVI.	LLLLVROYTE		
		150	160	170	180	190	200	JJJ 11(12
							200	
55			100	110	120	130	140	
	orf33.pep	VEMLAW	LPSKLGFPVI	PDARSVIEGRI	LNGNTADARA	WSGLLVXSTA		
		111111	:				1 11111	-
	orf33a	VEMT.AW	.PAKI.GFPVI	PDARAVIEGRI	NCNTADARA	WSGLLVGSTA	CVCTT.DDIT	ハいハリハマレ
	022000	210	220	230	240	250	260	MAVCK
60			220	200	240	230	200	
00	orf33a	ILXXTS	ENGLDLEKX	XXXXIRRWQ1	NKITDADTRR	ETVSAVSPKI	VLNDAPKWAY	VMLETE
		270	280	290	300	310	320	

## The complete length ORF33a nucleotide sequence <SEQ ID 201> is:

	• •	•		•			
	1	ATGTTGAATC	CATCCCGAAA	ACTGGTTGAG	CTGGTCCGTA	<b>ምምም</b> ርር አልር አ	
	51	AGGCGGCTTT	ATTTTCAGCG	GCGATCCCGT	GCAGGCGACG	GAGGCTTTGC	
_	101	GCCGCGTGGA	CGGCAGTACG	GAGGAAAAA	TCATCCGTCG	GGCGAAGATG	•
5	151	ATCGACAGGA	ACCGTATGCT	GCGGGAGACG	TTGGAACGTG	TGCGTGCGGG	:
	201	GTCGTTCTGG	TTGTGGGTGG	CGGCGGCGAC	GTTTGCGTTT	NTTACCGNTT	
	251	TTTCAGTTAC	TTATCTTCTA	ATGGACAATC	AGGGTCTGAA	TTTCTTTTTG	
	301	GTTTTGGCGG	GCGTGNTGGG	CATGAATACG	CTGATGCTGG	CAGTATGGTT	
10	351 401	GGCAATGTTG	TTCCTGCGCG	TGAAAGTGGG	GCGTTTTTTC	AGCAGTCCGG	•
10	451	TATECEGACE	ACTECCECCON	ACCUTGTCA	ATCAGGCGGT CGTTGGAAAA	GTTGCGGCTG	
	501	GTCGCACAGC	CTCTCCCTCT	CCACCCTCCT	CGGAATGCTG	CHCHCCCHAR	
	551	TGTTGCTGCT	TTTGGTGCGG	CAATATACGT	TCAACTGGGA	AACCACCCTC	
	601	TTGGGCGATT	CGTCTTCGGT	ACGGCTGGTG	GAAATGTTGG	CATGGCTGCC	•
15	651	TGCGAAACTG	GGTTTTCCCG	TGCCTGATGC	GCGGGCGGTC	ATCGAAGGTC	•
	701	GTCTGAACGG	CAATATTGCC	GATGCGCGGG	CTTGGTCGGG	GCTGCTGGTC	
	751	GGCAGTATCG	CCTGCTACGG	CATCCTGCCG	CGCCTCTTGG	CTTGGGCGGT	
	801	ATGCAAAATC	CTTNTGNAAA	CAAGCGAAAA	CGGCTTGGAT	TTGGAAAAGC	
20	851	NCNNNNNTCN	NNCGNTCATC	CGCCGCTGGC	AGAACAAAAT	CACCGATGCG	
20	901	GATACGCGTC	GGGAAACCGT	GTCCGCCGTT	TCGCCGAAAA	TCGTCTTGAA	
	951	CGATGCGCCG	AAATGGGCGG	TCATGCTGGA	GACCGAATGG	CAGGACGGCG	
	1001	AATGGTTCGA	GGGCAGGCTG	GCGCAGGAAT	GGCTGGATAA	GGGCGTTGCC	
	1051 1101	CCCCCAACCC	COUNTRACTOR	CGCGCTGGAG	ACAGAGCTGA	AGCAGAAACC	
25	1151	TCTTCCCCCA	CTTATUGGCG	CTTTTCCCAAC	AACTGTGCCC CGGCGCAGGG	GACCGCGGCG	
23	1201	GTGCANCTTT	TEECEEDACA	CITICGGAAG	GACGACCTTT	CCCAAAACC	
	1251	GGAACATTGG	CGTAACGCGC	TGACCGAATG	CGGCGCGCG	TECCTECANO	
	1301	CCGACAGAGC	GGCGCAGGAA	GGCCGTCTGA	AAACCAACGA	CCGCACTTGA	
						•	
	This encodes a p	rotein having	g amino acid	sequence <	SEQ ID 202	>:	
	•	*		_	-		
30	1	MLNPSRKLVE	LVRILEEGGF	IFSGDPVQAT	EALRRVDGST	EEKIIRRAKM	
٠	51	IDRNRMLRET	LERVRAGS <u>FW</u>	LWVAAATFAF	XTXFSVTYLL	MDNQGLNFFL	
	101	VLAGVXGMNT	<u>LMLAV</u> WLAML	FLRVKVGRFF	SSPATWFRGK	DPVNQAVLRL	ě
	151	YADEWRXPSV	RWKIGATSHS	LWLCTLLGML	VSVLLLLLVR	QYTFNWESTL	
35	201	LGDSSSVRLV	EMLAWLPAKL	GFPVPDARAV	IEGRLNGNIA	DARAWSG <u>LLV</u>	
33	251 301	DTDDETUCAL	RTTAMAACKI	LXXTSENGLD	LEKXXXXXXI	RRWQNKITDA	
	351	DIKKEIVSAV	SEVIATIONS	KWAVMLETEW	QDGEWFEGRL DRGVLRQIVR	AQEWLDKGVA	
	401	VXI.LAFOGI.S	DDI.SEKI.EHW	PNALTECCAA	WLEPDRAAQE	CRIMMINDA	
						GKTVINDKI.	
	ORF33a and OR	F33-1 show	94.1% ident	ity in 444 aa	overlap:		
					•		
40		•	10	20	30 4	10 50	60
	orf33a.pep	MLNPSP	KLVELVRILE	GGFIFSGDPV	<b>ATEALRRVDGS</b>	TEEKIIRRAK	MIDRNRMLRET
•		11111	111111111:1				
	orf33-1	MLNPSP	KLVELVRILDE	GGFIFSGDPV	DATEALRRVDGS	TEEKI IRRAEN	MIDRNRMLRET
45			10	20	30 4	10 50	60
43			70	00			
	orf33a.per	. I EDUDA	70	80	90 10	00 110	120
	Office of the control		GSEWLWVAAA]	THEATARSVI	YLLMDNQGLNFE	LVLAGVXGMN	LMLAVWLAML
	orf33-1	LERVRA	ווויוווו דממעעעענענער האדא א	וווו וווון עייניסיבים מיים מיים מיים	LLMDNQGLNF		NY NY DATES - 2007
50			70	80	90 10		
					10	,,,	120
	•	•	130		150 16	50 170	180
	orf33a.pep		GRFFSSPATWE	RGKDPVNQAVI	RLYADEWRXPS	VRWKIGATSHS	SIWICTII GMI
		111111	11111111111			111111111111	
55	orf33-1	FLRVKV	GRFFSSPATWE	RGKDPVNQAVI	RLYADEWROPS	VRWKIGATSHS	SLWLCTLLGML
			130		.50 16		
	•						
			190	200 2	210 22	230	240
60	orf33a.per		LLVKQYTFNWE	STLLGDSSSVR	RLVEMLAWLPAK	LGFPVPDARAV	/IEGRLNGNIA
UU	orf33-1	11111	111111111111	::::      			
	01133-1	∏انا ۱۸ ۵ ۸	190		AVEMLAWLPSK		
			170	200 2	210 22	230	240
			250	260 2	270 28	.290	300
65	orf33a.per	DARAWS			KILXXTSENGL	DLEKXXXXXX	עטעגאאטטאן פטעזאאטטאן
							XT T D C

_	1	5	8-

	522 1	DADAMSC	TINGSTACYG	TT.PRLLAWV	VCKILLKTSE	NGLDLEKPYY	QAVIRRWQNK	ITDA
	orf33-1	DAMA	250	260	270	280	290	300
			310	320	330	340	350	360
5	orf33a.pep	DTRRET	1111111111111		111111111	EGRLAQEWLD 		1111
	orf33-1	DTRRET	VSAVSPKIIL 310	NDAPKWAVML	ETEWQDGEWF 330	EGRLAQEWLD	KGVATNREQV. 350	AALE 360
10			370	380	390	400	410	420
10	orf33a.pep	111111	LIBERTALE		111111111	GGAVVXLLAE	11111111	FLIT
	orf33-1	TELKQK	PAQLLIGVRA 370	QTVPDRGVLR 380	QIVRLSEAAQ 390	GGAVVQLLAE 400	QGLSDDLSEK 410	420
15			430	440	450			
	orf33a.pep	RNALTE	CGAAWLEPDR	AAQEGRLKTN 	IDRTX		•	
	orf33-1	RNALAE	CGAAWLEPDR	AAQEGRLKDQ 440	ΣX			
20			430	440				

## Homology with a predicted ORF from N.gonorrhoeae

ORF33 shows 91.6% identity over a 143aa overlap with a predicted ORF (ORF33.ng) from N. gonorrhoeae:

25	orf33.pep	LFLRVKVGRFFSSPATWFRXKDPVNQAVLR			
	orf33ng	${\tt LMDNQGLNFFLVLAGVLGMNTLMLAVWLATLFLRVKVGRFFSSPATWFRGKGPVNQAVLR}$	100 		
30	orf33.pep	LYXDEWRXTSVRWKIXATSHSLWLCTLLGMLVSVLLLLLVRQYTFNWESTLLSNAASVRA	90		
	orf33ng	LYADOWROPSVRWKIGATAHSLWLCTLLGMLVSVLLLLLVRQYTFNWESTLLSNAASVRA	160		
	orf33.pep	VEMLAWLPSKLGFPVPDARSVIEGRLNGNIADARAWSGLLVXSIACXGILPRL	143		
35	orf33ng		220		

An ORF33ng nucleotide sequence <SEQ ID 203> was predicted to encode a protein having amino acid sequence <SEQ ID 204>:

```
1 MIDRDRMLRD TLERVRAGSF WLWVVVASMM FTAGFSGTYL LMDNQGLNFF
                                    LVLAGVLGMN TLMLAVWLAT LFLRVKVGRF FSSPATWFRG KGPVNQAVLR
                            51 LVLAGVLGMN TLMLAVWLAT LFLRVKVGRF FSSPATWFRG
101 LYADQWRQPS VRWKIGATAH SLWLCTLLGM LVSVLLLLLV RQYTFNWEST
151 LLSNAASVRA VEMLAWLPSK LGFPVPDARA VIEGRLNGNI ADARAWSGLL
151 LLSNAASVRA VEMLAWLPSK LGFPVPDARA VIEGRLNGNI ADARAWSGLL
152 LLSNAASVRA VEMLAWLPSK LGFPVPDARA VIEGRLNGNI RRWQNKITD
40
                            201 VGSIVCYGIL PRLLAWVVCK ILLKTSENGL DLEKTYYQAV IRRWQNKITD
                            251 ADTRRETVSA VSPKIVLNDA PKWALMLETE WQDGQWFEGR LAQEWLDKGV
301 AANREQVAAL ETELKQKPAQ LLIGVRAQTV PDRGVLRQIV RLSEAAQGGA
                            351 VVQLLAEQGL SDDLSEKLEH WRNALTECGA AWLEPDRVAQ EGRLKDQ*
45
```

# Further sequence analysis revealed the following DNA sequence <SEQ ID 205>:

50 1 1 2 2 2 3	51 agggggtTTT .01 gccgcgtgga .51 atcgACAGGg .01 gtcgtTctgG .51 TTTCAGgcac .6TTTTggcgG	CATCCCgaAA attttcagcg cggcAGTACG accgtatgtt TTATGGGTGG ttatCttCTG GAGTGTtggG TTCCTGCGCG	gcgatcctgt GAggAaaaa gcgggACaCg TggtggCAtC ATGGACaatC CATGaatacG TGAAAGTGGG	gcaggcgacg tcttccgtcg TtggaacGTG gATGATGTtt AGGGGCtGAA ctgATGCTGG ACGGTTTTTC	gaggettege GGCGGAGAtg TGCGTGCggg aCCGCCGGAT TtTCTTTTTA CAGTATGGtt AGCAGTCCGG
60	GACGTGGTT  GACGTGGTT  GACGTGGTC  GACGCACAG  GACCACAG  GACCACACAG  GACCACACAG  GACCACACACA	TCGGGGCAAA C AGTGGCGGCA C TTGTGGCTCT TTTGGTGCGGG C CGCTTCGGT C GGTTTCCCTG C CAATATTGCC TCTGCTACGG	GGCCCTGTAA ACCTTCGGTA GCACGCTGCT CAATATACGT ACGCGCGGTG TCCCCGATGC GATGCGCGGG	ATCAGGCGGT CGATGGAAAA CGGAATGCTG TCAACTGGGA GAAATGTTGG GCGGGCGGTC CTTGGTCGGG	TTGCGGCTG TAGGCGCAAC GTGTCGGTAT AAGCACGCTG CATGGCTGCC ATCGAAGGTC GCTGCTGGTC

5	851 0 901 0 951 0 1001 A 1051 0 1101 0 1151 1 1201 0	TGTAAAATC CTTTT CTATTATCA GGCGG ATACGCGTC GGGAA GATGCGCCG AAATG ATGGTTCGA GGCCA CCCAATCGGG AACAG GCGCAACTG CTTAT GCTGCGGCA GATTG GGAACATTGG CGTAA TGACAGGGT GGCGC	TCATC CGCC ACCGT GTCC GGCGC TCAT GGCTG GCGC GTTGC CGCG CGGCG TACG TGCGG CTTT GAACA GGGG CGCGC TGAC	GCTGGC AGAAC GCCGTT TCGCC GCTGGA GACCG AGGAAT GGCTG CTGGAG ACAGA CGCCCA AACTG CGGAAG CGGCG CTTTCA GACGA CGAATG CGGCG	AAAAT CACCGAT GAAAA TCGTCTT AGTGG CAGGACG GATAA GGGCGTT GCTGA AGCAGAA TGCCG GACCGGG CAGGG CGGCGCC CCTTT CGGAAAA CCGGCG TGGCTTC	rgcg rgaa ggcc rgcc lacc ggcg ggtg lgct
	This encodes a pro-	otein having amir	o acid sequ	ience <seq i<="" th=""><th>D 206; ORF33</th><th>ing-1&gt;:</th></seq>	D 206; ORF33	ing-1>:
15 20	51 1 101 V 151 V 201 I 251 C 301 I	MLNPSRKLVE LVRII DRDRMLRDT LERVR VLAGVLGMNT LMLAV VADQWRQPSV RWKIG SSIVCYGILP RLLAW DTRRETVSAV SPKIV ANREQVAALE TELKQ	AGSFW LWVV WLATL FLRV ATAHS LWLC LPSKL GFPV VVCKI LLKT LNDAP KWAL	VASMMF TAGFS KVGRFF SSPAT TLLGML VSVLI PDARAV IEGRI SENGLD LEKTY MLETEW QDGQW	GTYLL MDNQGLM WFRGK GPVNQAV LLLLVR QYTFNWE NGNIA DARAWSO YQAVI RRWQNKI FEGRL AQEWLDF	IF <u>FL</u> VLRL CSTL GLLV TDA KGVA
		QLLAEQGLS DDLSE				
•	ORF33ng-1 and C	ORF33-1 show 94	.6% identit	y in 446 aa o	verlap:	
25	orf33-1.pep	11111111111	1111::1111	111111111111111111111111111111111111111	111111111111111111111111111111111111111	50 60 RRAEMIDRNRMLRET              RAEMIDRDRMLRDT   50 60
30	orf33-1.pep	1011111111	: ::   .WVVVASMMFT	:        : AGFSGTYLLMDN		110 120 LGMNTLMLAVWLAML
35	orf33-1.pe	130 FLRVKVGRFFS				110 120 170 180 ATSHSLWLCTLLGML
40	orf33ng-1	130	140	150	160	ATAHSLWLCTLLGML 170 180
	orf33-1.pe					230 240 DARAVIEGRLNGNIA
45	orf33ng-1					DARAVIEGRLNGNIA 230 240
	orf33-1.pe	250 DARAWSGLLV	260 SIACYGILPR	270 RLLAWVVCKILLE	280 TSENGLDLEKPY	290 300 YQAVIRRWQNKITDA
50	orf33ng-1		:       SSIVCYGILPR 260	IIIIIIIIIIIIII LLAWVVCKILLE 270	TITITITIE TE STATE TO THE STATE	
55	orf33-1.pep	1111111111	111:11111	11:111111111	: 11111111111111	350 360 DKGVATNREQVAALE      :        DKGVAANREQVAALE 350 360
60	orf33-1.pe	370 TELKOKPAOLI	380 .TGVRAOTVPI	390	400	410 420 EQGLSDDLSEKLEHW
	orf33ng-1	1444141111	111111111111111111111111111111111111111			EQGLSDULSEKLEHW EQGLSDDLSEKLEHW 410 420
65	orf33-1.pe	430 P RNALAECGAAN				

orf33ng-1 RNALTECGAAWLEPDRVAQEGRLKDQX 430 440

Based on the presence of several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 25

The following partial DNA sequence was identified in N.meningitidis <SEQ ID 207>:

```
..CAGAAGAGTT TGTCGAGAAT TTCTTTATGG GGTTTGGGCG GCGTGTTTTT
                        CGGGGTGTCC GGTCTGGTAT GGTTTTCTTT GGGCGTTTCT TT.GAGTGCG
10
                 51
                        CCTGTTTTTC GGGTGTTTCT TTTCGGGGTT CGGGACGGGG GACGTTTGTG
                101
                        GGCAGTACGG GGGTTTCTTT GAGTGTGTTT TCAGCTTGTG TTCC.GGCGT
                 151
                        CGTCCGGCTG CCTGTCGGTT TGAGCTGTGT CGGCAGGTTG CG..GTTTGA
                 201
                        CCCGGTTTTT CTTGGGTGCG GCAGGGGACG TCATTCTCCT GCCGCTTTCG
                 251
                        TCTGTGCCGT CCGGCTGTGC GGGTTCGGAT GAGGCGGCGT GGTGGTGTTC
                 301
15
                        GGGTTGGGCG GCATCTTGT\underline{\mathbf{T}} CCGACTACGC CGTTTGGCAG CCAGAATTCG
                 351
                        GTTTCGCGGG GGCTGTCGGT GTGTTGCGGT TCGGCTTGAA GGGTTTTGTC
                 401
                 451
                        GTCC..
```

This corresponds to the amino acid sequence <SEQ ID 208; ORF34>:

```
20 1 ..QKSLSRISLW GLGGVFFGVS GLVWFSLGVS XECACFSGVS FRGSGRGTFV
51 GSTGVSLSVF SACVXGVVRL PVGLSCVGRL XXLTRFFLGA AGDVILLPLS
101 SVPSGCAGSD EAAWWCSGWA ASCPTTPFGS QNSVSRGLSV CCGSA*RVLS
151 S..
```

Further work revealed the complete nucleotide sequence <SEQ ID 209>:

						mcccmccccm
25	1	ATGATGATGC		GCTTCCTTGG		TGCCTGCCGT
	51	GCCGGGTCAG		001101212		TTGGGCGGCG
	101	TGTTTTTCGG	GGTGTCCGGT	11001111001		CGTTTCTTTG
	151	GGCTGCGCCT	GTTTTTCGGG		CGGGGTTCGG	
	201	GTTTGTGGGC	AGTACGGGGG		TGTGTTTTCA	
30	251	CGGCGTCGTC	CGGCTGCCTG	10001110	CTGTGTCGGC	
30	301	TTGACCCGGT	TTTTCTTGGG	200000	0.100001010	CGCTGCCGCT
	351	TTCGTCTGTG	CCGTCCGGCT	0100000	001110110000	GCGTGGTGGT
	401	GTTCGGGTTG	GGCGGCATCT	101000		
	451	TCGGTTTCGC	GGGGGCTGTC	GGTGTGTTGC	GGTTCGGCTT	GAAGGGTTTT
35	501	GTCGCCGTTC	GGGTTGAATG	1001011-	GCCTATTGCC	AATGCGCCGA
33	551	TGGCGGCGAT	ACAGATGAGC	AATACGGCGC	GTATCAGGAG	TTTGGGGGTC
	601	AGCCTGAAGG	GTTTGTTCGG	TTTTTTTGCC	ATTTTGATTG	TGCTTTTGGG
	651	GTGTCGGGCA		AAGGCGGTTC		GCCGAGTCAG
	701	CGTTGGACGT	AGTTTTGGTA	GAGGGTGATG	ACTTTTTGTA	CGCCGACGGT
40	751	GGTGCTGACT	TTTTGGGTAA	TCTGCGCCTG	TTCTTCGGGG	
40	801	CCATAACGTA	GGTTACGTTG	CCGTAGGTAA	CGATTTTGAC	
	851	GTGGCGGGGC	TGATGCCCAA	CAGCGTGGCG	CGGACTTTGG	
	901	AGTGTCGCCG	GCGATGTCGC	CGGCAGTGCG	CGGCAGGGAG	GCGACGGTAA
	951	TATAGTTGTA		GCGGCCTGTT	CGGAACGTGC	AATCTGACCG
45	1001	ACGAACTGTT		GGTGGCGACT	TGTCCGAGCA	GCAGCAGGTG
43	1051	GCGGTTGTAG		AGATTTGGGG	CGTGTAGCCT	TTGGTTTGGT
	1101	TGTTTTGGCG			TTTCGATACG	CAACGCCATA
	1151	ACGTTGTCGT			CGGCGGTCGA	CGGCGGATTT
	1201	CGCGCCGACG				CCGCTAAGGG
50	1251			ATCAGGGTGC		CGGTTTGGGT
50		TTCATCGGGT		TTGGGCGTTT		
	1301	CATGCCGTCT				
	1351	CAIGCCGTCI	. un			

This corresponds to the amino acid sequence <SEQ ID 210; ORF34-1>:

1	MMMPFIMLPW IAGV	PAVPGQ NRLSRISLWG LGGVE	FGVSG LVWFSLGVSL
55 51 101 151	GCACFSGVSF RGSG	RGTFVG STGVSLSVFS ACVP LPLSSV PSGCAGSDEA AWWCS RVLSPF GLNVLTMPIA NAPM	ASSGCL SV*AVSAGCG SGWAAS CPTTPFGSQN

	201	SLKGLFGFFA	ILIVLLGCRA	MPSEGGSDGI	AESALDVVLV	EGDDFLYADG
	251	GADFLGNLRL	FFGGEDAHNV	GYVAVGNDFD	ARLCGGADAQ	QRGADFGCVP
•	301	SVAGDVAGSA	RQGGDGNIVV	HAFGGLFGTC	NLTDELFFAF	GGDLSEQQQV
	. 351	AVVADDGDLG	RVAFGLVVLA	QIGTGGGFDT	QRHNVVVGLR	AGGSAVDGGF
5	401	RADGGASDYC	ADAAAKGKAE	NGGNQGADGV	RFGFHRVLPF	LGVSDGIALR
	451	HAV*				

Computer analysis of this amino acid sequence gave the following results:

### Homology with a predicted ORF from N. meningitidis (strain A)

ORF34 shows 73.3% identity over a 161aa overlap with an ORF (ORF34a) from strain A of N.

### 10 meningitidis:

				10	20	30	
	orf34.pep		. Q	KSLSRISLWG:	LGGVFFGVSGL	VWFSLGVSXE	CAC
			1	1 111 1111	111 1111111	$\overline{1}11111\overline{1}\overline{1}\overline{1}\overline{1}$	TH
	orf34a	MMXPXIMLPWI	AGVPAVPGQ	KRLSRXSLWG	LGGXFFGVSGL	VWFSLGVSXS	LGVSXGCAC
15		10	20	30	40	50	60
					•		
		40	50	60	70 .	80	90
	orf34.pep	FSGVSFRGSGR	GTFVGSTGV	SLSVFSACVX	GVVRLPVGLSC	VGRLXX	-LTRFFLGA
	,	<u> </u>	11111111		1:: :1	::	111 1 11:
20	orf34a	FSGVSFRGSGR	GTFVGSTGV	SLSVFSACA-	PASSGC	LSVXAVSAGO	GLTRXFXGA
		70	80		90	100	110
							•
		100	110	120	130	140	150
	orf34.pep	AGDVILLPLSS	VPSGCAGSD	EAAWWCSGWA	ASCPTTPFGSQ	NSVSRGLSVC	CGSAXRVLS
25		111 11111	11111111:1	1 1 11111		11111111111	111: 1111
	orf34a	AGDGSPLPLSS	VPSGCAGAD	EEAXXCSGWA	ASCPTTPFGSQ	NSVSRGLSVC	CGSVWRVLS
	•	120	130	140	150	160	170
30	orf34.pep	S	•				
	orf34a	PFGXNVLTMP1	_				AMPSEGGSD
		180	190	200	210	220	230

### The complete length ORF34a nucleotide sequence <SEQ ID 211> is:

35		1	ATGATGATNC	CGTTNATAAT	GCTTCCTTGG	ATTGCGGGTG	TGCCTGCCGT
		51	GCCGGGTCAG	AAGAGGTTGT	CGAGAANTTC	TTTATGGGGT	TTAGGCGGCN
		101	TGTTTTTCGG	GGTGTCCGGT	TTGGTATGGT	TTTCTTTGGG	CGTTTCTNTT
		151	TCTTTGGGTG	TTTCTNTGGG	CTGTGCCTGT	TTTTCGGGTG	TTTCTTTTCG
		201	GGGTTCGGGA	CGGGGGACGT	TTGTGGGCAG	TACNGGGGTT	TCTTTGAGTG
40		251	TGTTTTCAGC	TTGTGCTCCG	GCGTCGTCCG	GCTGCCTGTC	GGTTTNAGCT
		301	GTGTCGGCAG	GTTGCGGTTT	GACCCGGNTT	TTCTTNGGTG	CGGCAGGGGA
		351	CGGCAGTCCG	CTGCCGCTTT	CGTCTGTGCC	GTCCGGCTGT	GCGGGTGCGG
		401	ATGAGGAGGC	GTNGTNGTGT	TCGGGTTGGG	CGGCATCTTG	TCCGACTACG
		451	CCGTTTGGCA	GCCAGAATTC	GGTTTCGCGG	GGGCTGTCGG	TGTGTTGCGG
45		501	TTCGGTNTGG	AGGGTTTTGT	CNCCGTTCGG	GTNGAATGTG	CTGACGATGC
		551	CTATTGCCAA	TGCGCCGATG	GCGGTGATAC	AGATGAGCAA	TACGGCGCGT
		601	ATCAGGAGTT	TGGGGGTCAG	CCTGAAGGGT	TTGTTCNGTT	TTTTTGCCAT
		651	TTTGATTGTG	CTTTTGGGGT	GTCGGGCAAT	GCCGTCTGAA	GGCGGTTCAG
		701	ACGGCATTGC	CGAGTCAGCG	TTGGACGTAG	TTTNGGTAGA	GGGTGATGAC
50		751	TTTTTGTACG	CCGACGGTGG	TGCTGACTTT	TTGGGTAATC	TGCGCCTGTT
		801	CTTCGGGGGT	GAGGATGCCC	ATAACGTAGG	TTACGTTGCC	GTAGGTAACG
•		851	ATTTTGACGC	GCGCCTGTGT	GGCGGGGCTG	ATGCCCAACA	GCGTGGCGCG
		901	GACTTTGGAT	GTGTTCCAAG	TGTCGCCGGC	GATGTCGCCG	GCAGTGCGCG
		951	GCAGGGAGGC	GACGGTAATG	TANTTGTACA	CGCCTTCGGC	GGCCTGTTCG
55		1001	GAACGTGCAA	TCTGACCGAC	GAACTGTTTC	TCGCCTTCGG	TGGCGACTTG
		1051	TCCGAGCAGC	AGCAGGTGGC	GGTTGTAGCC	GACAACGGAG	ATTTGGGGCG
	;	1101	TGTANCCTTT	GGTTTGGTTG	TTTTGGCGCA	GATAGGAGCG	GGCGGTGGTT
		1151	TCGATACGCA	GCGCCATTAC	GTTGTCGTCG	GTTNGCGCGC	CGGTGGTTCG
		1201	GCGGTCGACG	GCGGATTTCG	CGCCGACCGC	CGCGCCGCCG	ACGACTGCGC
60		1251	TGACGCAGCC	GCCGAGGGCA	AGGCTGAGGA	CGGCGGCAGT	CAGGGTGCGG
		1301	ACGGTGTGCG	GTTTGGGTTT	CATCGGGTGC	TTCCTTTCTT	GGGCGTTTCA
		1351	GACGGCATTG	CTTTGCGCCA	TGCCGTCTGA		

This encodes a protein having amino acid sequence <SEQ ID 212>:

	1	MMXPXIMLPW	IAGVPAVPGQ	KRLSRXSLWG	LGGXFFGVSG	LVWFSLGVSX
	51	SLGVSXGCAC	FSGVSFRGSG	RGTFVGSTGV	SLSVFSACAP	ASSGCLSVXA
	101	VSAGCGLTRX	FXGAAGDGSP	LPLSSVPSGC	AGADEEAXXC	SGWAASCPTT
5	151	PFGSQNSVSR	GLSVCCGSVW	RVLSPFGXNV	LTMPIANAPM	AVIQMSNTAR
	201	IRSLGVSLKG	LFXFFAILIV	LLGCRAMPSE	GGSDGIAESA	LDVVXVEGDD
	251	FLYADGGADF	LGNLRLFFGG	EDAHNVGYVA	VGNDFDARLC	GGADAQQRGA
	301	DFGCVPSVAG	DVAGSARQGG	DGNVXVHAFG	GLFGTCNLTD	ELFLAFGGDL
	351	SEQQQVAVVA	DNGDLGRVXF	GLVVLAQIGA	GGGFDTQRHY	VVVGXRAGGS
10	401	AVDGGFRADR	RAADDCADAA	AEGKAEDGGS	QGADGVRFGF	HRVLPFLGVS
	451	DGIALRHAV*				

### ORF34a and ORF34-1 show 91.3% identity in 459 aa overlap:

		•		_		
15	orf34a.pep	10 20 MMXPXIMLPWIAGVPAVPGQ	KRLSRXSLWGLG			
15	orf34-1		NRLSRISLWGLG			 GCAC
20	orf34a.pep	70 80 FSGVSFRGSGRGTFVGSTGV	SLSVFSACAPAS			
	orf34-1	FSGVSFRGSGRGTFVGSTGV		SGCLSVXAVSAG	CGLTRFFLG	
25	orf34a.pep	130 140 LPLSSVPSGCAGADEEAXXC	SGWAASCPTTPF			
••	orf34-1	LPLSSVPSGCAGSDEAAWWC 120 130		GSQNSVSRGLSV	CCGSAXRVL	
30	orf34a.pep	190 200 LTMPIANAPMAVIQMSNTAR		220 XFFAILIVLLGO	230 RAMPSEGGSI	240 DGIAESA
35	orf34-1			GFFAILIVLLGO	RAMPSEGGSI	
	orf34a.pep	250 260 LDVVXVEGDDFLYADGGADF	LGNLRLFFGGED			
40	orf34-1	LDVVLVEGDDFLYADGGADF		AHNVGYVAVGND	FDARLCGGA	
45	orf34a.pep	310 320 DFGCVPSVAGDVAGSARQGG	DGNVXVHAFGGL			
	orf34-1	DFGCVPSVAGDVAGSARQGG 300 310	DGNIVVHAFGGL 320			QQVAVVA 350
50	orf34a.pep	370 380 DNGDLGRVXFGLVVLAQIGA	GGGFDTQRHYVV			
	orf34-1	DDGDLGRVAFGLVVLAQIGI 360 370	GGGFDTQRHNVV	VGLRAGGSAVDO	GFRADGGASI	
55	orf34a.pep	430 440 AEGKAEDGGSQGADGVRFGE	HRVLPFLGVSDG			
	orf34-1	AKGKAENGGNQGADGVRFGE	HRVLPFLGVSDG	IALRHAVX		
60	TT 1 1/2	420 430	440	450		

### Homology with a predicted ORF from N.gonorrhoeae

ORF34 shows 77.6% identity over a 161aa overlap with a predicted ORF (ORF34.ng) from N. gonorrhoeae:

	orf34ng		60
5	orf34.pep	FSGVSFRGSGRGTFVGSTGVSLSVFSACVXGVVRLPVGLSCVGRLXXLTRFFLGA	90
_	orf34ng	FSGVSFRGSGWGAFVGSTGVSLSVFSACVPVPVNESAARAASEGRGLTRFFLGA	114
	orf34.pep	AGDVILLPLSSVPSGCAGSDEAAWWCSGWAASCPTTPFGSQNSVSRGLSVCCGSAXRVLS	150
10	orf34ng	AGDGSPLPLSSVPSGCAGSDEAAWWCSGWAASCPTAPFGSQNSVSRGLSVCCGSVWRVLS	174
	orf34.pep	s	175
	orf34ng	PFGLNVLTMPTANAPMAVIQMSNTARIRSLGVSLKGLFGFFAILIVLLGCRAMPSEGGSD	234
15	The complete le	ngth ORF34ng nucleotide sequence <seq 213="" id=""> is:</seq>	
20	1 51 101 151 201 251 301	ATGATGATGC CGTTCATAAT GCTTCCTTGG ATTGCGGGTG TGCCTGCCGT GCCGGGTCAA AAGAGGTTGT CGAGAATCTC TTTATGGGGT TTGGCCGGCG TGTTTTTCGG GGTGTCCGGT TTGGTATGGT TTTCTTTGGG CGTTTCTTTT TCTTTGGGTG TTTCTTTGGG CTGCGCCTGT TTTTCGGGTG TTTCTTTTCG GGGTTCGGGA TGGGGGGCGT TTGTGGGCAG TACGGGGGTT TCTTTGAGTG TGTTTTCAGC TTGTGTTCCG GTGCCGGTTT ACGAATCGGC TGCCCGGGCC GCATCCGAAG GGCGCGGTTT GACCCGGTTT TTCTTGGGTG CGGCAGGGGA	
25	351 401 451 501 551	CCGTTTGGCA GCCAGAATTC GGTTTCGCGG GGGCTGTCGG TGTGTTGCGG TTCGGTTTGG AGGGTTTTGT CGCCGTTCGG GTTGAATGTG CTGACGATGC CTACTGCCAA TGCGCCGATG GCGGTGATAC AGATGAGCAA TACGGCGCGT	
30	601 651 701 751 801	TTTTTGTACG CCGAcggTGG TGCTGACTTT TTGGGTAATC TGCGCCTGTT	
. •	851 901	ATTTTGACGC GCGCCTGTGT AGCGGGGCTG ATGCCCAGCA GcgtgGCGCG	
35	951 1001 1051 1101 1151	GCAGGGAGGC GACGGTAATG TAGTTGTATA CGCCTTCGGC GGCCTGTTCG GAACGTGCAA TCTGACCGAC GAACTGTTTT TCGCCTTCGG TGGCGACTTG TCCGAGCAGC AGCAGGTGGC GGTTGTAGCC GACGACGGAG ATTTGGGGCG TGTAGCCTTT GGTTTGGTTG TTTTGGCGCA GGTAGGAACG GGCGGTGGTT TCGATACGCA ACGCCATAAC GTtgtCATCG GTTtqcqcqc CGGTGGTTcq	
40	1201 1251 1301 1351	TGAAGCAGCC GCCGAGGGCA AGGCTGAGGA CGGCGGCAAT CAGGGTGCGG ACGGTGTGTG GTTTGGGTTT CATCGGGGAC TTCCTTTCTT GGGCGTTTCA GACGGCATTG CTTTGCGCCA TGCCGTCTGA	
	This encodes a p	protein having amino acid sequence <seq 214="" id="">:</seq>	
45	1 51 101 151		
50	201 251 301 351 401	IRSLGVSLKG LFGFFAILIV LLGCRAMPSE GGSDGIAESA LDVVLVEGND FLYADGGADF LGNLRLFFGG EDAHNVGYIA VGNDFDARLC SGADAQQRGA DFGRVPSVAG DVARSARQGG DGNVVVYAFG GLFGTCNLTD ELFFAFGGDL SEQQQVAVVA DDGDLGRVAF GLVVLAQVGT GGGFDTQRHN VVIGLRAGGS AVDDGFCADG GPADDCAEAA AEGKAEDGGN QGADGVWFGF HRGLPFLGVS DGIALRHAV*	
55	•	ORF34-1 show 90.0% identity in 459 aa overlap:	
60	orf34-1.p	10 20 30 40 4 50  PEP MMMPFIMLPWIAGVPAVPGQNRLSRISLWGLGGVFFGVSGLVWFSLGVSLGC	11.
65	orf34-1.p	60 70 80 90 100 110  Pep FSGVSFRGSGRGTFVGSTGVSLSVFSACVPASSGCLSVXAVSAGCGLTRFFLGAAGDG:	11
,	3		

-164-

			70	80	90	100	110	120
5	orf34-1.pep orf34ng	111111		1111111	140 GWAASCPTTPE            GWAASCPTAPE 150	1111111111	111111: 11	11111111
10	orf34-1.pep	1111 [1		11111111	200 RSLGVSLKGLE           RSLGVSLKGLE 210	1111111111	1111111111	11111111
15	orf34-1.pep	1111111	:     EGNDFLY	 ADGGADFL	260 GNLRLFFGGED           GNLRLFFGGED	:     AHNVGYIAVG		 ADAQQRGA
20		300	250	260 310	270 320	280 330	290 340	300 350
	orf34-1.pep	111 111		111111	GNIVVHAFGGL   :  :     GNVVVYAFGGI	1111111111	1111111111	HILLIH
25	,	360	310	320	330	340	350	360
30	orf34-1.pep orf34ng	DDGDLGF		VLAQIGTG	380 GGFDTQRHNVV            GGFDTQRHNVV 390	:1111111111		:1 11:11
35	orf34-1.pep orf34ng	1:1111:	:	111 1111	440 RVLPFLGVSDG            RGLPFLGVSDG 450	450 GIALRHAVX	-23	

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from N.meningitidis and N.gonorrhoeae, and their epitopes, could be 40 useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 26

The following partial DNA sequence was identified in N.meningitidis <SEQ ID 215>:

	1	ATGAAAACCT	TCTTCAAAAC	CCTTTCCGCC	GCCGCACTCG	CGCTCATCCT
45	51	CGCCGCCTGC	GGATT.CAAA	AAGACAGCGC	GCCCGCCGCA	TCCGCTTCTG
1						TCGGCACGAC
1				TGGTCAAAGA		
2	201	AGAAAAAAGG	CTACACCGTC	AAACTGGTCG	AGTTTACCGA	CTATGTACGC
2				GGGCGAGTTG		

- 50 This corresponds to the amino acid sequence <SEO ID 216; ORF4>:
  - 1 MKTFFKTLSA AALALILAAC G.QKDSAPAA SASAAADNGA AKKEIVFGTT 51 VGDFGDMVKE QIQAELEKKG YTVKLVEFTD YVRPNLALAE GEL

Further sequence analysis revealed the complete nucleotide sequence <SEQ ID 217>:

```
ATGAAAACCT TCTTCAAAAC CCTTTCCGCC GCCGCACTCG CGCTCATCCT
55
                51 CGCCGCCTGC GGCGGTCAAA AAGACAGCGC GCCCGCCGCA TCCGCTTCTG
               101 CCGCCGCCGA CAACGGCGCG GCGAAAAAAG AAATCGTCTT CGGCACGACC
               151 GTCGGCGACT TCGGCGATAT GGTCAAAGAA CAAATCCAAG CCGAGCTGGA
               201 GAAAAAAGGC TACACCGTCA AACTGGTCGA GTTTACCGAC TATGTACGCC
```

```
251
                    CGAATCTGGC ATTGGCTGAG GGCGAGTTGG ACATCAACGT CTTCCAACAC
                     AAACCCTATC TTGACGACTT CAAAAAGAA CACAATCTGG ACATCACCGA
                301
                    AGTCTTCCAA GTGCCGACCG CGCCTTTGGG ACTGTACCCG GGCAAGCTGA
               351
                     AATCGCTGGA AGAAGTCAAA GACGGCAGCA CCGTATCCGC GCCCAACGAC
               401
 5
                451
                     CCGTCCAACT TCGCCCGCGT CTTGGTGATG CTCGACGAAC TGGGTTGGAT
                     CAAACTCAAA GACGGCATCA ATCCGTTGAC CGCATCCAAA GCGGACATCG
                501
                551
                     CCGAGAACCT GAAAAACATC AAAATCGTCG AGCTTGAAGC CGCGCAACTG
                     CCGCGTAGCC GCGCCGACGT GGATTTTGCC GTCGTCAACG GCAACTACGC
                601
                651
                     CATAAGCAGC GGCATGAAGC TGACCGAAGC CCTGTTCCAA GAACCGAGCT
10
                701
                     TTGCCTATGT CAACTGGTCT GCCGTCAAAA CCGCCGACAA AGACAGCCAA
                751
                     TGGCTTAAAG ACGTAACCGA GGCCTATAAC TCCGACGCGT TCAAAGCCTA
                801
                     CGCGCACAAA CGCTTCGAGG GCTACAAATC CCCTGCCGCA TGGAATGAAG
               851
                    GCGCAGCCAA ATAA
     This corresponds to the amino acid sequence <SEQ ID 218; ORF4-1>:
15
                     MKTFFKTLSA AALALILAAC GGQKDSAPAA SASAAADNGA AKKEIVFGTT
                     VGDFGDMVKE QIQAELEKKG YTVKLVEFTD YVRPNLALAE GELDINVFQH
```

101 KPYLDDFKKE HNLDITEVFQ VPTAPLGLYP GKLKSLEEVK DGSTVSAPND PSNFARVLVM LDELGWIKLK DGINPLTASK ADIAENLKNI KIVELEAAQL 151

201 PRSRADVDFA VVNGNYAISS GMKLTEALFQ EPSFAYVNWS AVKTADKDSQ 251 WLKDVTEAYN SDAFKAYAHK RFEGYKSPAA WNEGAAK\*

Computer analysis of this amino acid sequence gave the following results:

### Homology with a predicted ORF from N. meningitidis (strain A)

ORF4 shows 93.5% identity over a 93aa overlap with an ORF (ORF4a) from strain A of N. meningitidis:

```
25
                            10
                                     20
                                               30
                                                        40
                                                                  50
                     MKTFFKTLSAAALALILAACG-QKDSAPAASASAAADNGAAKKEIVFGTTVGDFGDMVKE
         orf4.pep
                     orf4a
                     MKTFFKTLSAAALALILAACGGQKDSAPAASASAAADNGAAXKEIVFGTTVGDFGDMVKE
                                              30
                                                       40
                                                                          60
30
                    60
                                      80
         orf4.pep
                     QIQAELEKKGYTVKLVEFTDYVRPNLALAEGEL
                      11 11111111111111 11111 11111
         orf4a
                     XIQPELEKKGYTVKLVEXTDYVRXNLALAEGELDINVXQHXXYLDDXKKXHNLDITXVXQ
35
                                     80
                                              90
                                                       100
                                                                110
                     VPTAPLGLYPGKLKSLXXVKXGSTVSAPNDPXXFXRVLVMLDELGXIKLKDXIXXXXXXX
         orf4a
                           130
                                    140
                                             150
```

The complete length ORF4a nucleotide sequence <SEQ ID 219> is:

40 1 ATGAAAACCT TCTTCAAAAC CCTTTCCGCC GCCGCACTCG CGCTCATCCT 51 CGCCGCCTGC GGCGGTCAAA AAGATAGCGC GCCCGCCGCA TCCGCTTCTC 101 CCGCCGCGA CAACGGCGCG GCGAANAAAG AAATCGTCTT CGGCACGACC 151 GTCGGCGACT TCGGCGATAT GGTCAAAGAA CANATCCAAC CCGAGCTGGA 201 GAAAAAAGGC TACACCGTCA AACTGGTCGA GTNTACCGAC TATGTGCGCC 45 CGAATCTGGC ATTGGCTGAG GGCGAGTTGG ACATCAACGT CTTNCAACAC 301 ANACNCTATC TTGACCACTN CAAAAAANAA CACAATCTGG ACATCACCNI 351 AGTCTTNCAA GTGCCGACCG CGCCTTTGGG ACTGTACCCG GGCAAGCTGA 401 AATCGCTGGA NNAAGTCAAA GANGCAGCA CCGTATCCGC GCCCAACGAC 451 CCGTNNNACT TCGNCCGCGT CTTGGTGAGCAC TGGGTTNGAT 451 CCGTNNNACT TCGNCCGCGT CTTGGTGACGAC TGGGTTNGAT 551 NNNNANNNNN NNNNNNNNNN NNNNNNNNNNN NNNNNN	
101 CCGCCGCGA CAACGGCGCG GCGAANAAG AAATCGTCTT CGGCACGACGACGACGACGACGACGACGACGACGACGACG	40
101 CCGCCGCGA CAACGGCGCG GCGAANAAAG AAATCGTCTT CGGCACGACCG 151 GTCGGCGACT TCGGCGATAT GGTCAAAGAA CANATCCAAC CCGAGCTGGA 201 GAAAAAAGGC TACACCGTCA AACTGGTCGA GTNTACCGAC TATGTGCGCC 45 CGAATCTGGC ATTGGCTGAG GGCGAGTTGG ACATCAACGT CTTNCAACAC 301 ANACNCTATC TTGACGACTN CAAAAAANAA CACAATCTGG ACATCACCNN 351 AGTCTTNCAA GTGCCGACCG CGCCTTTGGG ACTGTACCCG GGCAAGCTGA 401 AATCGCTGGA NNAAGTCAAA GANGGCAGCA CCGTATCCGC GCCCAACGAC 451 CCGTNNNACT TCGNCCGCGT CTTGGTGATG CTCGACGAAC TGGGTTNGAT 50 501 CAAACTCAAA GACNGCATCA NNNNGNNGNN NNNANCNANA NNNGANANNN 551 NNNNANNNNT NNNNNNNNN NNNNNCNNCG NNNNNNNNNNNNN	
45 251 CGAATCTGCC ATTGGCTGAG GCGAGTTGG ACATCAACGT CTTNCAACAC 301 ANACNCTATC TTGACGACTO CAAAAAANAA CACAATCTGG ACATCACCNI 351 AGTCTTNCAA GTGCCGACCG CGCCTTTGGG ACTGACCCCG GCCAACGAC 401 AATCGCTGGA NNAAGTCAAA GANGGCAGCA CCGTATCCGC GCCCAACGAC 451 CCGTNNNACT TCGNCCGCGT CTTGGTGATG CTCGACGAAC TGGGTTNGAT 50 501 CAAACTCAAA GACNGCATCA NNNNGNNGNN NNNAANCNANA NNNGANANNN 551 NNNNANNNT NNNNNNNNN NNNNNNNNNN NNNNNNNN	·
45 251 CGAATCTGGC ATTGGCTGAG GGCGAGTTGG ACATCAACGT CTTNCAACAC 301 ANACNCTATC TTGACGACTN CAAAAAANAA CACAATCTGG ACATCACCNN 351 AGTCTTNCAA GTGCCGACCG CGCCTTTGGG ACTGTACCCG GGCAAGCTGA 401 AATCGCTGGA NNAAGTCAAA GANGGCAGCA CCGTATCCGC GCCCAACGAC 451 CCGTNNNACT TCGNCCGCGT CTTGGTGATG CTCGACGAAC TGGGTTNGAT 501 CAAACTCAAA GACNGCATCA NNNNGNNGNN NNNAANCNANA NNNGANANN 551 NNNNANNNT NNNNNNNNN NNNNNCNNCG NNNNNNNNNN NNNNNNNN	:
45 251 CGAATCTGGC ATTGGCTGAG GGCGAGTTGG ACATCAACGT CTTNCAACAC 301 ANACNCTATC TTGACGACTN CAAAAAANAA CACAATCTGG ACATCACCNN 351 AGTCTTNCAA GTGCCGACCG CGCCTTTGGG ACTGTACCCG GGCAAGCTGA 401 AATCGCTGGA NNAAGTCAAA GANGGCAGCA CCGTATCCGC GCCCAACGAC 451 CCGTNNNACT TCGNCCGCGT CTTGGTGATG CTCGACGAAC TGGGTTNGAT 501 CAAACTCAAA GACNGCATCA NNNNGNNGNN NNNAANCNANA NNNGANANN 551 NNNNANNNT NNNNNNNNN NNNNNCNNCG NNNNNNNNNN NNNNNNNN	
351 AGTCTTNCAA GTGCCGACCG CGCCTTTGGG ACTGTACCCG GGCAAGCTGA 401 AATCGCTGGA NNAAGTCAAA GANGGCAGCA CCGTATCCGC GCCCAACGAC 451 CCGTNNACT TCGNCCGCGT CTTGGTGATG CTCGACGAAC TGGGTTNGAT 501 CAAACTCAAA GACNGCATCA NNNNGNNGNN NNNANCNANA NNNGANANNN 551 NNNNANNNNT NNNNNNNNNN NNNNNCNNCG NNNNNNNNNNNN	45
401 AATCGCTGGA NNAAGTCAAA GANGGCAGCA CCGTATCCGC GCCCAACGAC 451 CCGTNNNACT TCGNCCGCGT CTTGGTGATG CTCGACGAAC TGGGTTNGAT 501 CAAACTCAAA GACNGCATCA NNNNGNNGNN NNNANCNANA NNNGANANNN 551 NNNNANNNNT NNNNNNNNNN NNNNNCNNCG NNNNNNNNNN NNNNNNNN	;
50 CCGTNNNACT TCGNCCGCGT CTTGGTGATG CTCGACGAAC TGGGTTNGAT CAAACTCAAA GACNGCATCA NNNNGNNGNN NNNANCNANA NNNGANANNN NNNNANNNNT NNNNNNNNNN NNNNNCNNCG NNNNNNNNNNN NNNNNNNN	;
50 501 CAAACTCAAA GACNGCATCA NNNNGNNGNN NNNANCNANA NNNGANANNN 551 NNNNANNNNT NNNNNNNNNN NNNNNCNNCG NNNNNNNNNN NNNNNNNN	
50 501 CAAACTCAAA GACNGCATCA NNNNGNNGNN NNNANCNANA NNNGANANNN 551 NNNNANNNNT NNNNNNNNNN NNNNNCNNCG NNNNNNNNNN NNNNNNNN	•
551 NNNNANNNT NNNNNNNN NNNNNCNNCG NNNNNNNANN NNNNNNNNN	50
601 NCGNNTNNNN NNGCNNNNNT NNANNNTNNN NNCNNCNNNN NNNNNTNNN	!
	•
651 NANNANNAGC GGCATGAAGC TGACCGAAGC CCTGTTCCAA GAACCGAGC	1
701 TTGCCTATGT CAACTGGTCT GCCGTCAAAA CCGCCGACAA AGACAGCCA	
55 751 TGGCTTAAAG ACGTAACCGA GGCCTATAAC TCCGACGCGT TCAAAGCCTI	55
801 CGCGCACAAA CGCTTCGAGG GCTACAAATC CCCTGCCGCA TGGAATGAA	
851 GCGCAGCCAA ATAA	!

This is predicted to encode a protein having amino acid sequence <SEQ ID 220>:

MKTFFKTLSA AALALILAAC GGQKDSAPAA SASAAADNGA AXKEIVFGTT

```
VGDFGDMVKE XIQPELEKKG YTVKLVEXTD YVRXNLALAE GELDINVXQH XXYLDDXKKX HNLDITXVXQ VPTAPLGLYP GKLKSLXXVK XGSTVSAPND
                  101
                        PXXFXRVLVM LDELGXIKLK DXIXXXXXXX XXXXXXXXX XXXXXXXXX
                  151
                        XXXXXXXXX XXXXXXXXS GMKLTEALFQ EPSFAYVNWS AVKTADKDSQ
                  201
5
                        WLKDVTEAYN SDAFKAYAHK RFEGYKSPAA WNEGAAK*
                  251
```

#### A leader peptide is underlined.

### Further analysis of these strain A sequences revealed the complete DNA sequence <SEQ ID 221>:

	1	ATGAAAACCT	TCTTCAAAAC	CCTTTCCGCC	GCCGCACTCG	CGCTCATCCT
	51	CGCCGCCTGC	GGCGGTCAAA	AAGATAGCGC	GCCCGCCGCA	TCCGCTTCTG
10	101	CCGCCGCCGA	CAACGGCGCG	GCGAAAAAAG	AAATCGTCTT	CGGCACGACC
	151	GTCGGCGACT	TCGGCGATAT	GGTCAAAGAA	CAAATCCAAC	CCGAGCTGGA
	201	GAAAAAAGGC	TACACCGTCA	AACTGGTCGA	GTTTACCGAC	TATGTGCGCC
	251	CGAATCTGGC	ATTGGCTGAG	GGCGAGTTGG	ACATCAACGT	CTTCCAACAC
	301	AAACCCTATC	TTGACGACTT	CAAAAAAGAA	CACAATCTGG	ACATCACCGA
15	351	AGTCTTCCAA	GTGCCGACCG	CGCCTTTGGG	ACTGTACCCG	GGCAAGCTGA
••	401	AATCGCTGGA	AGAAGTCAAA	GACGGCAGCA	CCGTATCCGC	GCCCAACGAC
	451	CCGTCCAACT	TCGCCCGCGT	CTTGGTGATG	CTCGACGAAC	TGGGTTGGAT
	501	CAAACTCAAA	GACGGCATCA	ATCCGCTGAC	CGCATCCAAA	GCGGACATTG
	551	CCGAAAACCT	GAAAAACATC	AAAATCGTCG	AGCTTGAAGC	CGCGCAACTG
20	601	CCGCGTAGCC	GCGCCGACGT	GGATTTTGCC	GTCGTCAACG	GCAACTACGC
	651	CATAAGCAGC	GGCATGAAGC	TGACCGAAGC	CCTGTTCCAA	GAACCGAGCT
	701	TTGCCTATGT	CAACTGGTCT	GCCGTCAAAA	CCGCCGACAA	AGACAGCCAA
	751	TGGCTTAAAG	ACGTAACCGA	GGCCTATAAC	TCCGACGCGT	TCAAAGCCTA
	801	CGCGCACAAA	CGCTTCGAGG	GCTACAAATC	CCCTGCCGCA	TGGAATGAAG
25	851	GCGCAGCCAA	ATAA	•		

## This encodes a protein having amino acid sequence <SEQ ID 222; ORF4a-1>:

30	51 101 151	VGDFGDMVKE KPYLDDFKKE PSNFARVLVM	QIQPELEKKG HNLDITEVFQ LDELGWIKLK	YTVKLVEFTD VPTAPLGLYP DGINPLTASK	YVRPNLALAE GKLKSLEEVK ADIAENLKNI	AKKEIVFGTT GELDINVFQH DGSTVSAPND KIVELEAAQL AVKTADKDSO
2	201	PRSRADVDFA	<b>VVNGNYAISS</b>		EPSFAYVNWS	AVKTADKDSQ

#### ORF4a-1 and ORF4-1 show 99.7% identity in 287 aa overlap:

		10	20	30 40	50	60
35	orf4a-1	MKTFFKTLSAAALALIL	AACGGQKDSAP	AASASAAADNGAAI	KKEIVFGTTVGDFGDM	IVKE
<i>JJ</i>	-	1111111111111111111				
	orf4-1	MKTFFKTLSAAALALILA				
	0111 1	10		30 40	50	60
40		70	80	90 100	110	120
	orf4a-1	QIQPELEKKGYTVKLVE	FTDYVRPNLAL	AEGELDINVFQHK	PYLDDFKKEHNLDITE	VFQ
		111 111111111111			13 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	
	orf4-1	QIQAELEKKGYTVKLVE	FTDYVRPNLAL	AEGELDINVFOHK	PYLDDFKKEHNLDITE	VFQ
	-	70	80	90 100	110	120
45		•				
-1 <i>2</i>		130	140 1	.50 160	170	180
	orf4a-1	VPTAPLGLYPGKLKSLE	EVKDGSTVSAP	PNDPSNFARVLVML	DELGWIKLKDGINPL	<b>PASK</b>
		111111111111111111	111111111111		1111111111111	1111
	orf4-1	VPTAPLGLYPGKLKSLE	EVKDGSTVSAF	NDPSNFARVLVML	DELGWIKLKDGINPLY	rask
50		130	140 1	150 160	170	180
50						
		190	200 2	210 220	230	240
	orf4a-1	ADIAENLKNIKIVELEA	AOLPRSRADVI	FAVVNGNYAISSG	MKLTEALFOEPSFAY	/NWS
		1111111111111111				
55	orf4-1	ADIAENLKNIKIVELEA				
33	0111 1			210 220	230	240
		130				
		250	260 2	270 280		
	orf4a-1	AVKTADKDSQWLKDVTE	<b>AYNSDAFKAY</b>	AHKRFEGYKSPAAW	NEGAAKX	
60		1111111111111111111				
00	orf4-1	AVKTADKDSOWLKDVTE	AYNSDAFKAY	AHKRFEGYKSPAAW	NEGAAKX	
	0			200		

260

270

250

Homology with an outer membrane protein of Pasteurella haemolitica (accession q08869). ORF4 and this outer membrane protein show 33% aa identity in 91aa overlap:

	ORF4 and this out	er memorane pro	Mem silom 33	% aa idendiy	m 91aa ov	епар:	
	•	• • • • • • • • • • • • • • • • • • • •				20	
5	lip2.pasha	•		MNFKKLLG		ACKDEKAQAP	
,	ORF4			<b>NMKTFFKTLSA</b>	AALALILA	: :   ACGFKKTARPPHPL	
		110	120	130	140	150	
10		30 40		60	70	80	
10	lip2.pasha	-ATTAKTENKAPI	KVGVMTGPEAQM.  ::: :::		YGLDVELVQF  :  :	TEYTQPNAALHSKD	
	ORF4		VFGTTVGDFGDM	VKEQIQAELEK	KGYTVKLVEF	TDYVRPNLALAEGE	
_	•	160 170	180	190	200	210	
15	lip2.pasha	90 100		120	130	140 LKDGATVAIPNNAS	
	• •	1	QLVNDRGIRDAL	TONTENTIAL	·	TUDGAI VAI FNNAS	
	ORF4	L			•		
20	Homology with a p	oredicted ORF f	om <i>N.gonorr</i>	<u>hoeae</u>			
	ORF4 shows 93.6	% identity over	a 94aa over	lan with a n	redicted O	RF (ORF4 ng) fi	rom N
		, c				(Old 1.116) 1	
	gonorrhoeae:	•					
					10	20 30	٠
25	orf4nm.pep					LAACGXQKDSAPAA	1
	orf4ng			TSGENMKTFFK	TLSTASLALI	LAACGGQKDSAPAA	
		200	210	220	230	240 - 250	
30	orf4nm.pep	40	50	60	70	80 89 VEFTDYVRPNLALA	
50	• •	11:1:1:11111		111111111111	11111111111		
	orf4ng	SAAAPSADNGAAI 260	KEIVFGTTVGDF 270	GDMVKEQIQAE 280	LEKKGYTVKL 290	VEFTDYVRPNLALA 300 310	
35						300 310	
33	orf4nm.pep	90 EGEL					
	orf4ng	FGELDINVFOHK	YI.DDEKKEHNI.C	ם מידים היינו	T.CT. VDCKT KS	LEEVKDGSTVSAPN	
	011g	320	330	340	350	360 370	
40	The complete leng	gth ORF4ng nuc	eleotide seque	nce <seq i<="" th=""><th>D 223&gt; wa</th><th>s predicted to er</th><th>code a</th></seq>	D 223> wa	s predicted to er	code a
	protein having am	•				-	
	protein having and	mo acid scquenc		.4~.			
	1 M	KTFFKTLST ASLA	LILAAC GGQKDS	APAA SAAAPS	ADNG AAKKE	IVFGT	
	51 T 101 H	VGDFGDMVK EQIQA KPYLDDFKK EHNLA	ELEKK GYTVKI DITEAF QVPTAF	VEFT DYVRPN LGLY PGKLKS	LALA EGELD LEEV KDGST	INVFQ VSAPN	•
45	151 D	PSNFARALV MLNE	LGWIKL KDGINE	LTAS KADIAE	NLKN IKIVE	LEAAQ	
	251 Q	PRSRADVDF AVVNO WLKDVTEAY NSDA	FKAYAH KRFEGY	KYPA AWNEGA	AK*	AUKUS	
	Further analysis re	vealed the com	lete length Ol	RF4ng DNA	sequence <	SEO ID 225> to	be:
					_		<b>.</b>
50		tgAAAACCT TCTT( GCAGCCTGc ggCg					
	101 C	CCCTTCTGC CGAT	AACGgc gCgGCG	AAAA AAGAAA	tcgt ctTCG	GCACG	
		ccgtgggcg actt GAGAAAAAA Ggct					
	251 g	CCCGAATCT GGCA	TTGGCG GAGGGC	GAGT TGGACA	TCAA CGTCT	TCCAA	
55	301 C 351 C	ACAAACCCT ATCT GAAGCCTTC CAAG	rgacga tttcaa rgccga ccgccc	AAAA GAACAC CTTT GGGACT	AACC TGGAC	ATCAC CAAAC	
	401 T	GAAATCGCT GGAA	GAAGTC AAAGAC	GGCA GCACCG	TATC CGCGC	CCAac	
		ACCCGTCCA ACTT ATCAAACTC AAAG					
<b>60</b> .		CGCGGAAAA CCTG					

:11 1:11 111:: 1::: ::

orf4ng-1.pep

-168-

```
CTGCCGCGCA GCCGCGCGA CGTGGATTTT GCCGTCGTCA ACGGCAACTA
                   CGCCATAAGC AGCGGCATGA AGCTGACCGA AGCCCTGTTC CAAGAGCCGA
               651
                   GCTTTGCCTA TGTCAACTGG TCTGCCgtcA AAACCGCCGA CAAAGACAGC
               701
                   CAATGGCTTA AAGACGTAAC CGAGGCCTAT AACTCCGACG CGTTCAAAGC
               751
5
                   CTACGCGCAC AAACGCTTCG AGGGCTACAA ATACCCTGCC GCATGGAATG
               801
               851 AAGGCGCAGC CAAATAA
     This encodes a protein having amino acid sequence <SEQ ID 226; ORF4ng-1>:
                   MKTFFKTLSA AALALILAAC GGQKDSAPAA SAAAPSADNG AAKKEIVFGT
                    TVGDFGDMVK EQIQAELEKK GYTVKLVEFT DYVRPNLALA EGELDINVFQ
                51
                   HKPYLDDFKK EHNLDITEAF QVPTAPLGLY PGKLKSLEEV KDGSTVSAPN
10
               101
                   DPSNFARALV MLNELGWIKL KDGINPLTAS KADIAENLKN IKIVELEAAQ
               151
                    LPRSRADVDF AVVNGNYAIS SGMKLTEALF QEPSFAYVNW SAVKTADKDS
               201
                   QWLKDVTEAY NSDAFKAYAH KRFEGYKYPA AWNEGAAK*
               251
     This shows 97.6% identity in 288 as overlap with ORF4-1:
15
                                         20
                                                  30
                                                             40
                                                                      50
                        MKTFFKTLSAAALALILAACGGQKDSAPAASASA-AADNGAAKKEIVFGTTVGDFGDMVK
          orf4-1.pep
                        {\tt MKTFFKTLSAAALALILAACGGQKDSAPAASAAAPSADNGAAKKEIVFGTTVGDFGDMVK}
          orf4ng-1
                                10
                                         20
                                                   30
                                                            40
                                                                     50
20
                                                    90
                                                            100
                                          80
                        EQIQAELEKKGYTVKLVEFTDYVRPNLALAEGELDINVFQHKPYLDDFKKEHNLDITEVF
          orf4-1.pep
                         EQIQAELEKKGYTVKLVEFTDYVRPNLALAEGELDINVFQHKPYLDDFKKEHNLDITEAF
          orf4ng-1
25
                                         80
                                                   90
                                                           100
                                         140
                                                   150
                                                            160
                       120
                         QVPTAPLGLYPGKLKSLEEVKDGSTVSAPNDPSNFARVLVMLDELGWIKLKDGINPLTAS
          orf4-1.pep
                         QVPTAPLGLYPGKLKSLEEVKDGSTVSAPNDPSNFARALVMLNELGWIKLKDGINPLTAS
30
          orf4ng-1
                                        140
                                                  150
                                                           160
                                                                    170
                       180
                                190
                                         200
                                                   210
                                                            220
                                                                      230
                                                                              239
                         KADIAENLKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNW
          orf4-1.pep
35
                         KADIAENLKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNW
          orf4ng-1
                               190
                                        200
                                                  210
                                                           220
                                                                     230
                                                                              240
                                250
                                         260
                                                   270
                                                            280
                         SAVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEGYKSPAAWNEGAAKX
40
           orf4-1.pep
                         SAVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEGYKYPAAWNEGAAKX
           orf4ng-1
                               250
                                         260
                                                  270
                                                           280
      In addition, ORF4ng-1 shows significant homology with an outer membrane protein from the
 45
      database:
           ID
               LIP2 PASHA
                              STANDARD;
                                            PRT;
                                                  276 AA.
               Q08869;
           AC
                01-NOV-1995 (REL. 32, CREATED)
           DT
               01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
 50
           ידמ
               28.2 KD OUTER MEMBRANE PROTEIN PRECURSOR. . .
           DE
                      Init1:
                              279 Initn:
                                           416 Opt:
                                                    494
                                       36.0% identity in 275 aa overlap
           Smith-Waterman score: 494;
 55
                                                     30
                                                                        50
                                            20
                                 10
                         MKTFFKTLSAAAL--ALILAACGGQKDSAPAASAAAPSADNGAAKKEIVFGTTVGDFGDM
           orf4ng-1.pep
                             11 | ::|| || || |:|| :| :|| :||
                                                           :::| | |
                                                                        1::1::1
                           MNFKKLLGVALVSALALTACKDEKAQAPATTA---KTENKAPLK---VGVMTGPEAQM
           lip2 pasha
 60
                                  10
                                            20
                                                     30
                                                                 40
                                                                              50
                                  70
                                                             100
                                            80
                                                     90
                         VKEOIOAELEKKGYTVKLVEFTDYVRPNLALAEGELDINVFQHKPYLDDFKKEHNLDITE
```

11 | 1:11:11:1::11 | 1

	lip2_pasha	TEVAVKIA 60		VELVQFTE		KDLDANAFQT' 90	VPYLEQEVKD 100	RGYKLAI 110
_		120	130	140	150	160	170	
5	orf4ng-1.pep	AFQVPTAE			DGSTVSAPND:     :     :     :	PSNFARALVM        ::		
	lip2_pasha		PIAAYSK			ASNTARALLL 150		
10		180	190	200	210	220	23	0
_	orf4ng-1.pep	ASKADIAE		[VELEAAQL]		VVNGNYAISS		FQEPSFA : :
	lip2_pasha		ENPKNIKI BO	VQADTSLL	RMLDDVELA	VINNTYAGQA		
15							220	230
	orf4ng-1.pep	240 YVNWSAVI	250 CTADKDS( : :  :	260 WLKDVTEA'  : :::	YNSDAFKAYA	0 28 HKRFEGYKYP    :		
20	lip2_pasha		,	•	•	LKLFNGGVVK 270	GW	

Based on this analysis, including the homology with the outer membrane protein of *Pasteurella haemolitica*, and on the presence of a putative prokaryotic membrane lipoprotein lipid attachment site in the gonococcal protein, it was predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF4-1 (30kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figures 8A and 8B show, repsectively, the results of affinity purification of the His-fusion and GST-fusion proteins. Purified His-fusion protein was used to immunise mice, whose sera were used for ELISA (positive result), Western blot (Figure 8C), FACS analysis (Figure 8D), and a bactericidal assay (Figure 8E). These experiments confirm that ORF4-1 is a surface-exposed protein, and that it is a useful immunogen.

Figure 8F shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF4-1.

#### 35 **Example 27**

30

The following partial DNA sequence was identified in N. meningitidis <SEQ ID 227>:

```
CCTCGTCGTC CTCGGCATGC TCCAGTTTCA AGGGGCGATT TACTCCAAGG
                    CGGTGGAACG TATGCTCGGC ACGGTCATCG GGCTGGGCGC GGGTTTGGGC
               101
                    GTTTTATGGC TGAACCAGCA TTATTTCCAC GGCAACCTCC TCTTCTACCT
40
                    CACCGTCGGC ACGGCAAGCG CACTGGCCGG CTGGGCGGCG GTCGGCAAAA
               201
                    ACGGCTACGT CCCTmTGCTG GCAGGGCTGA CGATGTGTAT GCTCATCGGC
               251
                    GACAACGCA GCGAATGGCT CGACAGCGGA CTCATGCGCG CCATGAACGT
               301
                    CCTCATCGGC GYGGCCATCG CCATCGCCGC CGCCAAACTG CTGCCGCTGA
                    AATCCACACT GATGTGGCGT TTCATGCTTG CCGACAACCT GGCCGACTGC
               351
45
               401
                    AGCAAAATGA TTGCCGAAAT CAGCAACGGC AGGCGCATGA CCCGCGAACG
               451
                    CCTCGAGGAG AACATGGCGA AAATGCGCCA AATCAACGCA CGCATGGTCA
                    AAAGCCGCAG CCATCTCGCC GCCACATCGG GCGAAAGCTG CATCAGCCCC
               501
               551
                    GCCATGATGG AAGCCATGCA GCACGCCCAC CGTAAAATCG TCAACACCAC
               601
                    CGAGCTGCTC CTGACCACCG CCGCCAAGCT GCAATCTCCC AAACTCAACG
```

This corresponds to the amino acid sequence <SEQ ID 228; ORF8>:

```
1 .....PRRP RHAPVSRGDL LQGGGTYARH GHRAGRGFGR FMAEPALFPR
51 QPPLLPHRRH GKRTGRLGGG RQKRLRPXAG RADDVYAHRR QRQRMARQRT
10 101 HARHERPHRR GHRHRRQTA AAEIHTDVAF HACRQPGRLQ QNDCRNQQRQ
151 AHDPRTPRGE HGENAPNQRT HGQKPQPSRR HIGRKLHQPR HDGSHAARPP
201 XNRQHHRAAP DHRRQAAISQ TQRQRNPAAX PPLHTAPN......Q
251 TRPPHPHRHR HQPRTGSPRR TPPLPMAGLP LAQHRYASGN FRPRHPAATH
301 PPOMAGCPRT PTPAPKPA*
```

15 Computer analysis of this amino acid sequence gave the following results:

#### Sequence motifs

ORF8 is proline-rich and has a distribution of proline residues consistent with a surface localization. Furthermore the presence of an RGD motif may indicate a possible role in bacterial adhesion events.

20 Homology with a predicted ORF from N.gonorrhoeae

ORF8 shows 86.5% identity over a 312aa overlap with a predicted ORF (ORF8.ng) from N. gonorrhoeae:

```
orf8ng
                  1 MDRDDRLRRPRHAPVPRRDLLORGGTYARYGHRAGRGFGRFMAEPALFPR 50
                         25
                  1 .....PRRPRHAPVSRGDLLQGGGTYARHGHRAGRGFGRFMAEPALFPR 44
        orf8.pep
                 51 QPPLLPDHRHGKRTGRLGGGRQKRLRPYVGGADDVHAHRRQRQRMARQRP 100
        orf8ng
                    45 QPPLLPHRRHGKRTGRLGGGRQKRLRPXAGRADDVYAHRRQRQRMARQRT 94
        orf8.pep
30
                 101 DARDERPHRRHHCRRQTAAAEIHTDVAFHACRQPGRLQQNDCRNQQRQ 150
        orf8ng
                     95 HARHERPHRRGHRHRRRQTAAAEIHTDVAFHACRQPGRMQQNDCRNQQRQ 144
        orf8.pep
                 151 AYDARTFGAEYGQNAPNQRTHGQKPQPPRRHIGRKPHQPLHDGSHAARPP 200
35
        orf8ng
                           145 AHDPRTPRGEHGENAPNQRTHGQKPQPSRRHIGRKLHQPRHDGSHAARPP 194
        orf8.pep
                 201 QNRQHHRAAPDHRRQAAISQTQRQRNPAARPPLHTAPNRPATNRRPHQRQ 250
        orf8ng
40
                     195 XNRQHHRAAPDHRRQAAISQTQRQRNPAAXPPLHTAPN...........Q 244
        orf8.pep
                 251 TRPPHPHRHRHQPRTGSPRRTPPLPMAGFPLAQHQYASGNFRPRHPPATH 300
        orf8ng
                    45
                 245 TRPPHPHRHRHQPRTGSPRRTPPLPMAGLPLAQHRYASGNFRPRHPAATH 294
        orf8.pep
                 301 PPQMAGCPRTPTPAPKPA* 319
        orf8ng
                    1111111111111111111
                 295 PPQMAGCPRTPTPAPKPA* 313
        orf8.pep
```

The complete length ORF8ng nucleotide sequence <SEQ ID 229> is predicted to encode a protein having amino acid sequence <SEQ ID 230>:

- 1 MDRDDRLRRP RHAPVPRRDL LQRGGTYARY GHRAGRGFGR FMAEPALFPR
- 51 QPPLLPDHRH GKRTGRLGGG RQKRLRPYVG GADDVHAHRR QRQRMARQRP
- 101 DARDERPHRR RHRHCRRQTA AAEIHTDVAF HACRQPGRLQ QNDCRNQQRQ
- 55 151 AYDARTFGAE YGQNAPNQRT HGQKPQPPRR HIGRKPHQPL HDGSHAARPP

- 201 QNRQHHRAAP DHRRQAAISQ TQRQRNPAAR PPLHTAPNRP ATNRRPHQRQ
- 251 TRPPHPHRHR HQPRTGSPRR TPPLPMAGFP LAQHQYASGN FRPRHPPATH
- 301 PPOMAGCPRT PTPAPKPA\*

Based on the sequence motifs in these proteins, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 28

The following partial DNA sequence was identified in N. meningitidis <SEQ ID 231>:

```
..GAAATCAGCC TGCGGTCCGA CNACAGGCCG GTTTCCGTGN CGAAGCGGCG
10
                 51
                       GGATTCGGAA CGTTTTCTGC TGTTGGACGG CGGCAACAGC CGGCTCAAGT
                       GGGCGTGGGT GGAAAACGGC ACGTTCGCAA CCGTCGGTAG CGCGCCGTAC
                101
                       CGCGATTTGT CGCCTTTGGG CGCGGAGTGG GCGGAAAAGG CGGATGGAAA
                151
                       TGTCCGCATC GTCGGTTGCG CTGTGTGCGG AGAATTCAAA AAGGCACAAG
                201
                251
                       TGCAGGAACA GCTCGCCCGA AAAATCGAGT GGCTGCCGTC TTCCGCACAG
15
                       GCTTT.GGCA TACGCAACCA CTACCGCCAC CCCGAAGAAC ACGGTTCCGA
                301
                351
                       CCGCTGGTTC AACGCCTTGG GCAGCCGCCG CTTCAGCCGC AACGCCTGCG
                       TCGTCGTCAG TTGCGGCACG GCGGTAACGG TTGACGCGCT CACCGATGAC
                401
                451
                       GGACATTATC TCGGAGA.GG AACCATCATG CCCGGTTTCC ACCTGATGAA
                501
                       AGAATCGCTC GCCGTCCGAA CCGCCAACCT CAACCGGCAC GCCGGTAAGC
20
                551
                       GTTATCCTTT CCCGACCGG..
```

This corresponds to the amino acid sequence <SEQ ID 232; ORF61>:

```
1 ..EISLRSDXRP VSVXKRRDSE RFLLLDGGNS RLKWAWVENG TFATVGSAPY
51 RDLSPLGAEW AEKADGNVRI VGCAVCGEFK KAQVQEQLAR KIEWLPSSAQ
101 AXGIRNHYRH PEEHGSDRWF NALGSRRFSR NACVVVSCGT AVTVDALTDD
25 GHYLGXGTIM PGFHLMKESL AVRTANLNRH AGKRYPFPT..
```

Further work revealed the complete nucleotide sequence <SEQ ID 233>:

	1	ATGACGGTTT	TGAAGCTTTC	GCACTGGCGG	GTGTTGGCGG	AGCTTGCCGA
	51	CGGTTTGCCG	CAACACGTCT	CGCAACTGGC	GCGTATGGCG	GATATGAAGC
	101	CGCAGCAGCT	CAACGGTTTT	TGGCAGCAGA	TGCCGGCGCA	CATACGCGGG
30	151	CTGTTGCGCC	AACACGACGG	CTATTGGCGG	CTGGTGCGCC	CATTGGCGGT
	201	TTTCGATGCC	GAAGGTTTGC	GCGAGCTGGG	GGAAAGGTCG	GGTTTTCAGA
	251	CGGCATTGAA	GCACGAGTGC	GCGTCCAGCA	ACGACGAGAT	ACTGGAATTG
	301	GCGCGGATTG	CGCCGGACAA	GGCGCACAAA	ACCATATGCG	TGACCCACCT
	351				GAAGTGGTCG	
35	401	+ + +	-		TGTTTGACCG	
	451				GTGGCGTGTC	
	501				GTGGCCCAAT	
	551				TTGAAACGGT	
	601				ATCAATTTTG	
40	651				GCTGTTTCAG	
	701				TGGAAACGCT	
	751				GACGGATTTG	
	801				CGGCAAGGCG	
	851			-	CGGTTAAAGG	
45	901			_	GGCAAACAGA	
	951				GCCGGTTTCC	
	1001				ACGGCGGCAA	
	1051				GCAACCGTCG	
••	1101		•		GTGGGCGGAA	
50	1151				GCGGAGAATT	
	1201		-		GAGTGGCTGC	
	1251				CCACCCGAA	
	1301				GCCGCTTCAG	
	1351				ACGGTTGACG	
55	1401			-	GCCCGGTTTC	
	1451			_	TCAACCGGCA	
	1501			_	GCCGTCGCCA	
	1551					TTGAAAGAAA
	1601	AAACCGGGGC	GGGCAAGCCT	GTCGATGTCA	TCATTACCGG	CGGCGGCGCG

-172-

1651	GCAAAAGTTG	CCGAAGCCCT	GCCGCCTGCA	TTTTTGGCGG	AAAATACCGT
1701	GCGCGTGGCG	GACAACCTCG	TCATTTACGG	GTTGTTGAAC	ATGATTGCCG
1751	CCGAAGGCAG	GGAATATGAA	CATATTTAA		

This corresponds to the amino acid sequence <SEQ ID 234; ORF61-1>:

```
5
                 1 MTVLKLSHWR VLAELADGLP QHVSQLARMA DMKPQQLNGF WQQMPAHIRG
                    LLRQHDGYWR LVRPLAVFDA EGLRELGERS GFQTALKHEC ASSNDEILEL
                 51
               101
                    ARIAPDKAHK TICVTHLQSK GRGRQGRKWS HRLGECLMFS FGWVFDRPQY
                    ELGSLSPVAA VACRRALSRL GLDVQIKWPN DLVVGRDKLG GILIETVRTG
                201 GKTVAVVGIG INFVLPKEVE NAASVQSLFQ TASRRGNADA AVLLETLLVE
10
                251
                    LDAVLLQYAR DGFAPFVAEY QAANRDHGKA VLLLRDGETV FEGTVKGVDG
                301
                    QGVLHLETAE GKQTVVSGEI SLRSDDRPVS VPKRRDSERF LLLDGGNSRL
                351
                    KWAWVENGTF ATVGSAPYRD LSPLGAEWAE KADGNVRIVG CAVCGEFKKA
                    QVQEQLARKI EWLPSSAQAL GIRNHYRHPE EHGSDRWFNA LGSRRFSRNA
                401
                451
                    CVVVSCGTAV TVDALTDDGH YLGGTIMPGF HLMKESLAVR TANLNRHAGK
15
                501 RYPFPTTTGN AVASGMMDAV CGSVMMMHGR LKEKTGAGKP VDVIITGGGA
                    AKVAEALPPA FLAENTVRVA DNLVIYGLLN MIAAEGREYE HI*
```

Figure 9 shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF61-1. Further computer analysis of this amino acid sequence gave the following results:

Homology with the baf protein of B. pertussis (accession number U12020).

20 ORF61 and baf protein show 33% aa identity in 166aa overlap:

```
orf61 23
                    LLLDGGNSRLKWAWVE-NGTFATVGSAPYR----DLSPLGAEWAEKADGNVRIVGCAVCG 77
                     +L+D GNSRLK W + +
                                        Α
                                             AΡ
                                                     DL LG A
                                                                      R +G V G
                 3
          baf
                     ILIDSGNSRLKVGWFDPDAPQAAREPAPVAFDNLDLDALGRWLATLPRRPQRALGVNVAG 62
25
          orf61
                 78 EFKKAQVQEQLAR---KIEWLPSSAQAXGIRNHYRHPEEHGSDRW---FNALGSRRFSRN 131
                                   I WL + A G+RN YR+P++ G+DRW
                          + L
                                                                     L +
          baf
                 63
                     LARGEAIAATLRAGGCDIRWLRAQPLAMGLRNGYRNPDQLGADRWACMVGVLARQPSVHP 122
          orf61 132 ACVVVSCGTAVTVDALTDDGHYLGXGTIMPGFHLMKESLAVRTANL 177
30
                       +V S GTA T+D + D + G G I+PG +M+ +LA TA+L
          baf
                 123 PLLVASFGTATTLDTIGPDNVFPG-GLILPGPAMMRGALAYGTAHL 167
```

#### Homology with a predicted ORF from N.meningitidis (strain A)

ORF61 shows 97.4% identity over a 189aa overlap with an ORF (ORF61a) from strain A of N.

#### 35 meningitidis:

						10	20	30
	orf61.pep				EISLRSD	KRPVSVXKRRI	DSERFLLLD	€GNS
					111111		111111111	
40	orf6la		_	ILETAEGKQTV				€GNS
40		290	300	310	320	330	340	
			40	50	60	70	80	90
	orf61.pep			APYRDLSPLG				
		11111111	11111111	1111111111	111111:111			1111
45	orf6la	RLKWAWVE	NGTFATVGS	APYRDLSPLG	AEWAEKVDGN'	VRIVGCAVCG	EFKKAQVQE	<b>DLAR</b>
		350	360	370	380	390	400	
			^^	110	100	120	1.40	150
	661	_	00				140	150
50	orf61.pep	VIEWT522	AQAAGIRNE	IYRHPEEHGSD	KWINALGSKK	FSRNACVVVS	JGTAVTVDAL	7700
30	orf61a			IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII		IIIIIIIIIIII		1111
	Offora	410	420	430	440	450	460	7100
		410	420	430	440	450	400	
		1	60	170	180 1	89		
55	orf61.pep	GHYLGXGT	IMPGFHLME	ESLAVRTANL	NRHAGKRYPF	PT		
		11111 11				11		
	orf6la	GHYLG-GT	IMPGFHLM	<b>ESLAVRTANL</b>	NRHAGKRYPF	PTTTGNAVAS	GMMDAVCGS	VMMM
		470	480	490	500	510	520	
60	auf61a	HODE IN	~3 <i>~</i> 227317517	TMCCCS		.1001 F131 F131 F131 F13	TII/27 + 124	
60	orf6la	HGRLKERT	GAGKPVDV.	ITGGGAAKVA	EALPPAFLAE	NIVKVADNLV	THEFTULIA	AEGG

530 540 550 560 570 580 The complete length ORF61a nucleotide sequence <SEO ID 235> is: ATGACGGTTT TGAAGCCTTC GCACTGGCGG GTGTTGGCGG AGCTTGCCGA CGGTTTGCCG CAACACGTCT CGCAACTGGC GCGTATGGCG GATATGAAGC 51 5 CGCAGCAGCT CAACGGTTTT TGGCAGCAGA TGCCGGCGCA CATACGCGGG 101 CTGTTGCGCC AACACGACGG CTATTGGCGG CTGGTGCGCC CATTGGCGGT TTTCGATGCC GAAGGTTTGC GCGAGCTGGG GGAAAGGTCG GGTTTTCAGA 151 201 CGGCATTGAA GCACGAGTGC GCGTCCAGCA ACGACGAGAT ACTGGAATTG 251 GCGCGGATTG CGCCGGACAA GGCGCACAAA ACCATATGTG TGACCCACCT 301 GCAAAGTAAG GGCAGGGGGC GGCAGGGGCG GAAGTGGTCG CACCGTTTGG 10 351 GCGAGTGTCT GATGTTCAGT TTTGGCTGGG TGTTTGACCG GCCGCAGTAT 401 GAGTTGGGTT CGCTGTCGCC TGTTGCGGCA GTGGCGTGCC GGCGCGCCTT 451 GTCGCGTTTG GGTTTGAAAA CGCAAATCAA GTGGCCAAAC GATTTGGTCG 501 TCGGACGCGA CAAATTGGGC GGCATTCTGA TTGAAACGGT CAGGACGGGC 551 15 GGCAAAACGG TTGCCGTGGT CGGTATCGGC ATCAATTTCG TGCTGCCCAA 601 GGAAGTGGAA AACGCCGCTT CCGTGCAATC GCTGTTTCAG ACGGCATCGC 651 GGCGGGGAAA TGCCGATGCC GCCGTGTTGC TGGAAACGCT GTTGGCGGAA 701 CTTGATGCGG TGTTGTTGCA ATATGCGCGG GACGGATTTG CGCCTTTTGT 751 801 GGCGGAATAT CAGGCTGCCA ACCGCGACCA CGGCAAGGCG GTATTGCTGT 20 TGCGCGACGG CGAAACCGTG TTCGAAGGCA CGGTTAAAGG CGTGGACGGA 851 901 CAAGGCGTTC TGCACTTGGA AACGGCAGAG GGCAAACAGA CGGTCGTCAG CGGCGAAATC AGCCTGCGGT CCGACGACAG GCCGGTTTCC GTGCCGAAGC 951 GGCGGGATTC GGAACGTTTT CTGCTGTTGG ACGGCGGCAA CAGCCGGCTC 1001 AAGTGGGCGT GGGTGGAAAA CGGCACGTTC GCAACCGTCG GTAGCGCGCC 1051 25 GTACCGCGAT TTGTCGCCTT TGGGCGCGGA GTGGGCGGAA AAGGTGGATG 1101 1151 GAAATGTCCG CATCGTCGGT TGCGCCGTGT GCGGAGAATT CAAAAAGGCA 1201 CAAGTGCAGG AACAGCTCGC CCGAAAAATC GAGTGGCTGC CGTCTTCCGC ACAGGCTTTG GGCATACGCA ACCACTACCG CCACCCCGAA GAACACGGTT 1251 1301 CCGACCGCTG GTTCAACGCC TTGGGCAGCC GCCGCTTCAG CCGCAACGCC 30 1351 TGCGTCGTCG TCAGTTGCGG CACGGCGGTA ACGGTTGACG CGCTCACCGA TGACGGACAT TATCTCGGGG GAACCATCAT GCCCGGTTTC CACCTGATGA AAGAATCGCT CGCCGTCCGA ACCGCCAACC TCAACCGGCA CGCCGGTAAG 1401 1451 1501 CGTTATCCTT TCCCGACCAC AACGGGCAAT GCCGTCGCCA GCGGCATGAT GGATGCGGTT TGCGGCTCGG TTATGATGAT GCACGGGCGT TTGAAAGAAA 1551 35 1601 AAACCGGGGC GGGCAAGCCT GTCGATGTCA TCATTACCGG CGGCGGCGCG GCAAAAGTTG CCGAAGCCCT GCCGCCTGCA TTTTTGGCGG AAAATACCGT 1651 GCGCGTGGCG GACAACCTCG TCATTCACGG GCTGCTGAAC CTGATTGCCG 1701 CCGAAGGCGG GGAATCGGAA CATACTTAA 1751 This encodes a protein having amino acid sequence <SEQ ID 236>: 40 MTVLKPSHWR VLAELADGLP QHVSQLARMA DMKPQQLNGF WQQMPAHIRG LLRQHDGYWR LVRPLAVFDA EGLRELGERS GFQTALKHEC ASSNDEILEL 51 ARIAPDKAHK TICVTHLQSK GRGRQGRKWS HRLGECLMFS FGWVFDRPQY ELGSLSPVAA VACRRALSRL GLKTQIKWPN DLVVGRDKLG GILIETVRTG 151 GKTVAVVGIG INFVLPKEVE NAASVQSLFQ TASRRGNADA AVLLETLLAE 201 LDAVLLQYAR DGFAPFVAEY QAANRDHGKA VLLLRDGETV FEGTVKGVDG QGVLHLETAE GKQTVVSGEI SLRSDDRPVS VPKRRDSERF LLLDGGNSRL 45 251 301 KWAWVENGTF ATVGSAPYRD LSPLGAEWAE KVDGNVRIVG CAVCGEFKKA 351 QVQEQLARKI EWLPSSAQAL GIRNHYRHPE EHGSDRWFNA LGSRRFSRNA 401 CVVVSCGTAV TVDALTDDGH YLGGTIMPGF HLMKESLAVR TANLNRHAGK 451 50 RYPFPTTTGN AVASGMMDAV CGSVMMMHGR LKEKTGAGKP VDVIITGGGA 501 AKVAEALPPA FLAENTVRVA DNLVIHGLLN LIAAEGGESE HT\* ORF61a and ORF61-1 show 98.5% identity in 591 aa overlap: 10 20 30 40 MTVLKPSHWRVLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR orf6la.pep 55 MTVLKLSHWRVLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR orf61-1 10 20 30 40 RN 90 60 LVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK orf6la.pep LVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK orf61-1 70 80 90 100 110 65

140

150

160

170

180

130

WO 99/24578 PCT/IB98/01665

5	orf61a.pep orf61-1 orf61a.pep	GRGRQG:	RKWSHRLGECI            RKWSHRLGECI 130 190 DKLGGILIETY	 LMFSFGWVFDI 140 200 /RTGGKTVAV		  PVAAVACRRA  160     220  KEVENAASVQ	:   LSRLGLDVQI 170 230 SLFQTASRRG	IIII KWPN 180 240 NADA
10	orf61-1	DLVVGR	DKLGGILIET 190	/RTGGKTVAV 200	VGIGINFVLPI 210	KEVENAASVQ 220	SLFQTASRRG 230	NADA 240
15	orf61a.pep		250 LLAELDAVLLÇ   :       LLVELDAVLLÇ 250	HIIIIIIII	11111111		114111111	1111
20	orf61a.pep	111111	310 ETAEGKQTVVS            ETAEGKQTVVS 310	111111111	111111111			1111
25	orf61a.pep	11111	370 PYRDLSPLGAI           PYRDLSPLGAI 370			11111111	1111111111	LIH
30	orf61a.pep		430 RHPEEHGSDRI					
35	orf61-1	GIRNHY	RHPEEHGSDR 430 490	FNALGSRRF: 440 500	SRNACVVVSC 450 510	GTAVTVDALT 460 520	DDGHYLGGTI 470 530	MPGF 480 540
40	orf61a.pep	111111	LAVRTANLNRI            LAVRTANLNRI 490		111111111	11111111	1111111111	1111
45	orf61a.pep	111111	550 GGGAAKVAEAI            GGGAAKVAEAI 550			1111:1111	1-11	

### Homology with a predicted ORF from N.gonorrhoeae

ORF61 shows 94.2% identity over a 189aa overlap with a predicted ORF (ORF61.ng) from N.

### 50 gonorrhoeae:

	orf61.pep	EISLRSDXRPVSVXKRRDSERFLLLDGGNS	30
	orf61ng		211
55	orf61.pep	RLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGEFKKAQVQEQLAR	90
	orf61ng		271
60	orf61.pep	KIEWLPSSAQAXGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDD	150
	orf61ng	KIEWLPSSAQALGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDD	331
	orf61.pep	GHYLGXGTIMPGFHLMKESLAVRTANLNRHAGKRYPFPT	189
65	orf61ng	[	390

An ORF61ng nucleotide sequence <SEQ ID 237> was predicted to encode a protein having amino acid sequence <SEQ ID 238>:

```
1 MFSFGWAFDR PQYELGSLSP VAALACRRAL GCLGLETQIK WPNDLVVGRD
51 KLGGILIETV RAGGKTVAVV GIGINFVLPK EVENAASVQS LFQTASRRGN
101 ADAAVLLETL LAELGAVLEQ YAEEGFAPFL NEYETANRDH GKAVLLLRDG
151 ETVCEGTVKG VDGRGVLHLE TAEGEQTVVS GEISLRPDNR SVSVPKRPDS
201 ERFLLLEGGN SRLKWAWVEN GTFATVGSAP YRDLSPLGAE WAEKADGNVR
251 IVGCAVCGES KKAQVKEQLA RKIEWLPSSA QALGIRNHYR HPEEHGSDRW
301 FNALGSRRFS RNACVVVSCG TAVTVDALTD DGHYLGGTIM PGFHLMKESL
10 351 AVRTANLNRP AGKRYPFPTT TGNAVASGMM DAVCGSIMMM HGRLKEKNGA
401 GKPVDVIITG GGAAKVAEAL PPAFLAENTV RVADNLVIHG LINLIAAEGG
```

Further analysis revealed the complete gonococcal DNA sequence <SEQ ID 239> to be:

```
ATGACGGTTT TGAAGCCTTC GCATTGGCGG GTGTTGGCGG AGCTTGCCGA
                      CGGTTTGCCG CAACACGTAT CGCAATTGGC GCGTGAGGCG GACATGAAGC
15
                     CGCAGCAGCT CAACGGTTTT TGGCAGCAGA TGCCGGCGCA TATACGCGGG
                101
                      CTGTTGCGCC AACACGACGG CTATTGGCGG CTGGTGCGCC CCTTGGCGGT
                201
                     TTTCGATGCC GAAGGTTTGC GCGATCTGGG GGAAAGGTCG GGTTTTCAGA
                      CGGCATTGAA GCACGAGTGC GCGTCCAGCA ACGACGAGAT ACTGGAATTG
                251
20
                     GCGCGGATTG CGCCGGACAA GGCGCACAAA ACCATATGCG TGACCCACCT
                301
                      GCAAAGTAAG GGCAGGGGCC GGCAGGGGCC GAAGTGGTCC CACCGTTTGG
                351
                     GCGAGTGCCT GATGTTCAGT TTCGGCTGGG CGTTTGACCG GCCGCAGTAT
                401
                      GAGTTGGGTT CGCTGTCGCC TGTTGCGGCA CTTGCGTGCC GGCGCGCTTT
                451
                     GGGGTGTTTG GGTTTGGAAA CGCAAATCAA GTGGCCAAAC GATTTGGTCG
                501
25
                551
                     TCGGACGCGA CAAATTGGGC GGCATTCTGA TTGAAACAGT CAGGGCGGGC
                     GGTAAAACGG TTGCCGTGGT CGGTATCGGC ATCAATTTCG TGCTGCCCAA
                601
                     GGAAGTGGAA AACGCCGCTT CCGTGCAGTC GCTGTTTCAG ACGGCATCGC
GGCGGGGCAA TGCCGATGCC GCCGTATTGC TGGAAACATT GCTTGCGGAA
                651
                701
                      CTGGGCGCGG TGTTGGAACA ATATGCGGAA GAAGGGTTCG CGCCATTTTT
                 751
30
                801
                      AAATGAGTAT GAAACGGCCA ACCGCGACCA CGGCAAGGCG GTATTGCTGT
                     TGCGCGACGG CGAAACCGTG TGCGAAGGCA CGGTTAAAGG CGTGGACGGA CGAGGCGTTC TGCACTTGGA AACGGCAGaa ggcgaACAGa cggtcgtcag
                 851
                 901
                 951
                      cggcgaaaTC AGcctGCggc ccgacaacaG GTCGGtttcc gtgccgaagc
               1001
                      ggccggatTC GgaacgtTTT tTGCtgttgg aaggcgggaa cagccgGCTC
35
                      AAGTGGGCGT GggtggAAAa cggcacgttc gcaaccgtgg gcagcgcgCc
                1051
                      GTACCGCGAT TTGTCGCCTT TGGGCGCGGA GTGGGCGGAA AAGGCGGATG
               1101
               1151
                      GAAATGTCCG CATCGTCGGT TGCGCCGTGT GCGGAGAATC CAAAAAGGCA
                      CAAGTGAAGG AACAGCTCGC CCGAAAAATC GAGTGGCTGC CGTCTTCCGC
                1201
                1251
                      ACAGGCTTTG GGCATACGCA ACCACTACCG CCACCCCGAA GAACACGGTT
40
                      CCGACCGTTG GTTCAACGCC TTGGGCAGCC GCCGCTTCAG CCGCAACGCC
                1301
                1351
                      TGCGTCGTCG TCAGTTGCGG CACGGCGGTA ACGGTTGACG CGCTCACCGA
                      TGACGGACAT TATCTCGGCG GAACCATCAT GCCCGGCTTC CACCTGATGA
                1401
                1451
                      AAGAATCGCT CGCCGTCCGA ACCGCCAACC TCAACCGCCC CGCCGGCAAA
                1501
                      CGTTACCCTT TCCCGACCAC AACGGGCAAC GCCGTCGCAA GCGGCATGAT
45
                1551
                      GGACGCGGTT TGCGGCTCGA TAATGATGAT GCACGGCCGT TTGAAAGAAA
                      AAAACGGCGC GGGCAAGCCT GTCGATGTCA TCATTACCGG CGGCGCGCG
                1601
                1651
                      GCGAAAGTCG CCGAAGCCCT GCCGCCTGCA TTTTTGGCGG AAAATACCGT
                1701
                      GCGCGTGGCG GACAACCTCG TCATCCACGG GCTGCTGAAC CTGATTGCCG
                1751
                      CCGAAGGCGG GGAATCGGAA CACGCTTAA
```

50 This corresponds to the amino acid sequence <SEQ ID 240; ORF61ng-1>:

	. 1	MTVLKPSHWR	VLAELADGLP	QHVSQLAREA	DMKPQQLNGF	WQQMPAHIRG
	51	LLRQHDGYWR	LVRPLAVFDA	EGLRDLGERS	<b>GFQTALKHEC</b>	ASSNDEILEL
	101	ARIAPDKAHK	TICVTHLQSK	GRGRQGRKWS	HRLGECLMFS	FGWAFDRPQY
	151	ELGSLSPVAA	LACRRALGCL	GLETQIKWPN	DLVVGRDKLG	GILIETVRAG
55	201	GKTVAVVGIG	INFVLPKEVE	NAASVQSLFQ	TASRRGNADA	AVLLETLLAE
	251	LGAVLEQYAE	EGFAPFLNEY	ETANRDHGKA	VLLLRDGETV	CEGTVKGVDG
	301	RGVLHLETAE	GEQTVVSGEI	SLRPDNRSVS	VPKRPDSERF	LLLEGGNSRL
•	351	KWAWVENGTF	ATVGSAPYRD	LSPLGAEWAE	KADGNVRIVG	CAVCGESKKA
	401	QVKEQLARKI	EWLPSSAQAL	GIRNHYRHPE	EHGSDRWFNA	LGSRRFSRNA
60	451				HLMKESLAVR	
	501	RYPFPTTTGN	AVASGMMDAV	CGSIMMMHGR	LKEKNGAGKP	VDVIITGGGA
	551	AKVAEALPPA	FLAENTVRVA	DNLVIHGLLN	LIAAEGGESE	HA*

ORF61ng-1 and ORF61-1 show 93.9% identity in 591 aa overlap:

WO 99/24578 PC -176-

	orf61ng-1.pep	MTVLKPSHWRVLAELADGLPQHVSQLAREADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR 60	
	orf61-1	MTVLKLSHWRVLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR 60	
5	orf61ng-1.pep	LVRPLAVFDAEGLRDLGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK 120	
	orf61-1	LVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK 120	
10	orf6lng-1.pep	GRGRQGRKWSHRLGECLMFSFGWAFDRPQYELGSLSPVAALACRRALGCLGLETQIKWPN 180	
10	orf61-1	GRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLDVQIKWPN 180	,
	orf61ng-1.pep	DLVVGRDKLGGILIETVRAGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA 240	J
15	orf61-1	DLVVGRDKLGGILIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA 240	ı
	orf61ng-1.pep	AVLLETLLAELGAVLEQYAEEGFAPFLNEYETANRDHGKAVLLLRDGETVCEGTVKGVDG 300	)
20	orf61-1	AVLLETLLVELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDG 300	)
20	orf61ng-1.pep	RGVLHLETAEGEQTVVSGEISLRPDNRSVSVPKRPDSERFLLLEGGNSRLKWAWVENGTF 360	)
	orf61-1	QGVLHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTF 360	)
25	orf61ng-1.pep	ATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGESKKAQVKEQLARKIEWLPSSAQAL 420	)
	orf61-1	ATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQAL 420	)
20	orf6lng-1.pep	GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGF 480	)
30	orf61-1	GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGF 480	)
	orf6lng-1.pep	HLMKESLAVRTANLNRPAGKRYPFPTTTGNAVASGMMDAVCGSIMMMHGRLKEKNGAGKP 540	)
35	orf61-1	HLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKP 540	)
	orf6lng-1.pep	VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGGESEHAX 593	
40	orf61-1	VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIYGLLNMIAAEGREYEHIX 593	

Based on this analysis, including the homology with the baf protein of *B. pertussis* and the presence of a putative prokaryotic membrane lipoprotein lipid attachment site, it is predicted that these proteins from *N. meningitidis* and *N. gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### 45 Example 29

The following partial DNA sequence was identified in N.meningitidis <SEQ ID 241>:

	1	ATGTTTTACC	AAATCCTTGC	CCTGATTATC	TGGAGCAGCT	CGTTTATTGC
	51	CGCCAAATAT	GTCTATGGCG	GCATCGATCC	CGCATTGATG	GTCGGCGTGC
	101	GCCTGCTAAT	TGCCGCGCTG	CCTGCACTGC	CCGCCTGCCG	CCGTCATGTC
50	151	GGCAAGATTC	CGCGTGAGGA	ATGGAAGCCG	TTGCTGATTG	TGTCGTTCGT
• -	201	CAACTATGTG	CTGACCCTGC	TGCTTCAGTT	TGTCGGGTTG	AAATACACTT
	251	CCGCCGCCAG	CGCATCGGTC	ATTGTCGGAC	TCGAGCCGCT	GCTGATGGTG
	301		••••		GCGCGTGCCT	
	351				GCTGCTGATG	
55	401	CGGaAGAGGG	CGGCGaAGTC	GGCTGGTTCG	GCTGCCTGCT	GGTGTTGTTG
	451				CCGACGCAAA	
	501	ACGCATCGGC	GCACCGGCAT	TCACATCTGT	TTCCATTGCC	GCCGCATCGT
	551	TGATGTGCCT	GCCGTTTTCG	CTTGCTTTGG	CGCAAAGTTA	TACCGTGGAC
	601	TGGAGCGTCG	GGATGGTATT	GTCGCTGCTG	TATTTGGGTT	TGGGGTGC

This corresponds to the amino acid sequence <SEQ ID 242; ORF62>:

orf62.pep

	-111-
	1 NUMBER OF STREET
	1 MFYQILALII WSSSFIAAKY VYGGIDPALM VGVRLLIAAL PALPACRRHV 51 GKIPREEWKP LLIVSFVNYV LTLLLQFVGL KYTSAASASV IVGLEPLLMV
•	101 FVGHFFFNDK ARAYHWICGA AAFAGVALLM AGGAEEGGEV GWFGCLLVLL
	151 AGAGFCAAMR PTORLIARIG APAFTSVSIA AASLMCLPFS LALAQSYTVD
5	201 WSVGMVLSLL YLGLGC
_	T 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
	Further work revealed the complete nucleotide sequence <seq 243="" id="">:</seq>
	1 DECEMBER OF ANAMOUNDED COMPANIES DOCUMENTED DE COMPANIES
	1 ATGTTTTACC AAATCCTTGC CCTGATTATC TGGAGCAGCT CGTTTATTGC 51 CGCCAAATAT GTCTATGGCG GCATCGATCC CGCATTGATG GTCGGCGTGC
	101 GCCTGCTAAT TGCCGCGCTG CCTGCACTGC CCGCCTGCCG CCGTCATGTC
10	151 GGCAAGATTC CGCGTGAGGA ATGGAAGCCG TTGCTGATTG TGTCGTTCGT
	201 CAACTATGTG CTGACCCTGC TGCTTCAGTT TGTCGGGTTG AAATACACTT
	251 CCGCCGCCAG CGCATCGGTC ATTGTCGGAC TCGAGCCGCT GCTGATGGTG
	301 TTTGTCGGAC ACTTTTTCTT CAACGACAAA GCGCGTGCCT ACCACTGGAT
	351 ATGCGGCGCG GCGGCATTTG CCGGTGTCGC GCTGCTGATG GCGGGCGGTG
15	401 CGGAAGAGGG CGGCGAAGTC GGCTGGTTCG GCTGCCTGCT GGTGTTGTTG
	451 GCGGGCGCG GCTTTTGTGC CGCTATGCGT CCGACGCAAA GGCTGATTGC
	501 ACGCATCGGC GCACCGGCAT TCACATCTGT TTCCATTGCC GCCGCATCGT 551 TGATGTGCCT GCCGTTTTCG CTTGCTTTGG CGCAAAGTTA TACCGTGGAC
•	601 TGGAGGGTCG GGATGGTATT GTCGCTGCTG TATTTGGGTT TGGGGTGCGG
20	651 CTGGTACGCC TATTGGCTGT GGAACAAGGG GATGAGCCGT GTTCCTGCCA
	701 ATGTTTCGGG ACTGTTGATT TCGCTCGAAC CCGTCGTCGG CGTGCTGCTG
	751 GCGGTTTTGA TTTTGGGCGA ACACCTGTCG CCCGTGTCCG CCTTGGGCGT
	801 GTTTGTCGTC ATCGCCGCCA CCTTGGTTGC CGGCCGGCTG TCGCATCAAA
	851 AATAA
25	This corresponds to the amino acid sequence <seq 244;="" id="" orf62-1="">:</seq>
23	Time corresponds to the animo acre sequence - say is 211, ord on 1.
	1 MFYQILALII WSSSFIAAKY VYGGIDPALM VGVRLLIAAL PALPACRRHV
	51 GKIPREEWKP LLIVSFVNYV LTLLLQFVGL KYTSAASASV IVGLEPLLMV
	101 FVGHFFFNDK ARAYHWICGA AAFAGVALLM AGGAEEGGEV GWFGCLLVLL
30	151 AGAGFCAAMR PTQRLIARIG APAFTSVSIA AASLMCLPFS LALAQSYTVD
30	201 WSVGMVLSLL YLGLGCGWYA YWLWNKGMSR VPANVSGLLI SLEPVVGVLL 251 AVLILGEHLS PVSALGVFVV IAATLVAGRL SHOK*
	Computer analysis of this amino acid sequence gave the following results:
	Homology with hypothetical transmembrane protein HI0976 of H. influenzae (accession number Q57147)
	ORF62 and HI0976 show 50% aa identity in 114aa overlap:
25	
35	Orf62 1 MFYQILALIIWSSFIAAKYVYGGIDPALMVGVRXXXXXXXXXXXXCRRHVGKIPREEWKP 60
	M YQILAL+IWSSS I K Y +DP L+V VR R KI + K HI0976 1 MLYQILALLIWSSSLIVGKLTYSMMDPVLVVQVRLIIAMIIVMPLFLRRWKKIDKPMRKQ 60
	NIO 970 I MINI TIMBELLA DI VONDI ISMME LA VOVEDI IMMILI VME DE DERMENTO DE COMPANS DE CO
	Orf62 61 LLIVSFVNYVLTLLLQFVGLKYTSAASASVIVGLEPLLMVFVGHFFFNDKARAY 114
40	L ++F NY LLQF+GLKYTSA+SA ++GLEPLL+VFVGHFFF K +
	HI0976 61 LWWLAFFNYTAVFLLQFIGLKYTSASSAVTMIGLEPLLVVFVGHFFFKTKQNGF 114
	II 1
	Homology with a predicted ORF from N. meningitidis (strain A)
	ORF62 shows 99.5% identity over a 216aa overlap with an ORF (ORF62a) from strain A of N
	Old oz shows 33.376 Idolicky over a 210da overlap with all old (Old oza) Holli strain 71 of 17
45	meningitidis:
	10 20 30 40 50 60
	orf62.pep MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLLIAALPALPACRRHVGKIPREEWKP
50	orf62a MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLLIAALPALPACRRHVGKIPREEWKP
50	10 20 30 40 50 60
	70 80 90 100 110 120
	orf62.pep LLIVSFVNYVLTLLLQFVGLKYTSAASASVIVGLEPLLMVFVGHFFFNDKARAYHWICGA
55	orf62a LLIVSFVNYVLTLLLQFVGLKYTSAASASVIVGLEPLLMVFVGHFFFNDKARAYHWICGA
	70 80 90 100 110 120

130 140 150 160 170 180 AAFAGVALLMAGGAEEGGEVGWFGCLLVLLAGAGFCAAMRPTQRLIARIGAPAFTSVSIA

180

PCT/IB98/01665 WO 99/24578

-178-

	orf62a	AAFAGVALLMAGGAEEGGEVGWFGCLLVLLAGAGFCAAMRPTQRLIARIGAPAFTSVSIA  130 140 150 160 170 180
5		190 200 210
	orf62.pep	AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGLGC
	orf62a	AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGVGCSWYAYWLWNKGMSRVPANVSG <u>LLI</u>
10		190 200 210 220 230 240
10	orf62a	SLEPVVGVLLAVLI LGEHLSPVSVLGVFVVIAATLVAGRLSHQKX 250 260 270 280
	The complete leng	th ORF62a nucleotide sequence <seq 245="" id=""> is:</seq>
	<b>7</b> 70	
15		TGTTTTACC AAATCCTTGC CCTGATTATC TGGAGCAGCT CGTTTATTGC GCCAAATAT GTCTATGGCG GCATCGATCC CGCATTGATG GTCGGCGTGC
10	101 G	CCTGCTGAT TGCTGCGCTG CCTGCACTGC CCGCCTGCCG CCGTCATGTC
		GCAAGATTC CGCGTGAGGA ATGGAAGCCG TTGCTGATTG TGTCGTTCGT AACTATGTG CTGACCCTGC TACTTCAGTT TGTCGGGTTG AAATACACTT
		CGCCGCCAG CGCATCGGTC ATTGTCGGAC TCGAGCCACT GCTGATGGTG
20	301 T	TTGTCGGAC ACTTTTCTT CAACGACAAA GCGCGTGCCT ACCACTGGAT
		TGCGGCGCG GCGGCATTTG CCGGTGTCGC GCTGCTGATG GCGGGCGGTG
		CGGAAGAGGG CGGCGAAGTC GGCTGGTTCG GCTGCCTGCT GGTGTTGTTG CCGGGCGCGG GCTTTTGTGC CGCTATGCGT CCGACGCAAA GGCTGATTGC
	501 F	CGCATCGGC GCACCGGCAT TCACATCTGT TTCCATTGCC GCCGCATCGT
25		GATGTGCCT GCCGTTTTCG CTTGCTTTGG CGCAAAGTTA TACCGTGGAC
		GGAGCGTCG GAATGGTATT GTCGCTGCTG TATTTGGGCG TGGGGTGCAG TGGTACGCC TATTGGCTGT GGAACAAGGG GATGAGCCGT GTTCCTGCCA
		CGTTTCGGG ACTGTTGATT TCGCTCGAAC CCGTCGTCGG CGTGCTGCTG
20		CGGTTTTGA TTTTGGGCGA ACACCTGTCG CCCGTGTCCG TCTTGGGCGT
30		TTTGTCGTC ATCGCCGCCA CCTTGGTTGC CGGCCGGCTG TCGCATCAAA
		otein having amino acid sequence <seq 246="" id="">:</seq>
	Tins cheodes a pr	otoni naving mimo dola soquence 15DQ ID 240°.
		MFYQILALII WSSSFIAAKY VYGGIDPALM VGVRLLIAAL PALPACRRHV
25		SKIPREEWKP LLIVSFVNYV LTLLLQFVGL KYTSAASASV IVGLEPLLMV
35		TVGHFFFNDK ARAYHWICGA AAFAGVALLM AGGAEEGGEV GWFGCLLVLL AGAGFCAAMR PTQRLIARIG APAFTSVSIA AASLMCLPFS LALAQSYTVD
	201 V	SVGMVLSLL YLGVGCSWYA YWLWNKGMSR VPANVSGLLI SLEPVVGVLL
	251 2	AVLILGEHLS PVSVLGVFVV IAATLVAGRL SHOK*
	ORF62a and ORF	62-1 show 98.9% identity in 284 aa overlap:
40	orf62a.pep	MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLLIAALPALPACRRHVGKIPREEWKP 60
	550 1	
	orf62-1	MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLLIAALPALPACRRHVGKIPREEWKP 60
	orf62a.pep	LLIVSFVNYVLTLLLQFVGLKYTSAASASVIVGLEPLLMVFVGHFFFNDKARAYHWICGA 120
45		
	orf62-1	LLIVSFVNYVLTLLLQFVGLKYTSAASASVIVGLEPLLMVFVGHFFFNDKARAYHWICGA 120
	orf62a.pep	AAFAGVALLMAGGAEEGGEVGWFGCLLVLLAGAGFCAAMRPTQRLIARIGAPAFTSVSIA 180
50		
30	orf62-1	WALHOANTTIMOOWEEGGEAGMIGCTTATTWOMOLCWMIKEIÄKTTWKTGWENL12A2TH 100
	orf62a.pep	AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGVGCSWYAYWLWNKGMSRVPANVSGLLI 240
	orf62-1	
55	02102-1	
	orf62a.pep	SLEPVVGVLLAVLILGEHLSPVSVLGVFVVIAATLVAGRLSHQKX 285
	orf62-1	

#### Homology with a predicted ORF from N.gonorrhoeae 60

ORF62 shows 99.5% identity over a 216aa overlap with a predicted ORF (ORF62.ng) from N. gonorrhoeae:

		MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLLIAALPALPACRRHVGKIPREEWKP              MFYQILALIIWGSSFIAAKYVYGGIDPALMVGVRLLIAALPALPACRRHVGKIPREEWKP	60 60
5	orf62.pep	LLIVSFVNYVLTLLLQFVGLKYTSAASASVIVGLEPLLMVFVGHFFFNDKARAYHWICGA	120
	orf62ng		120
10	orf62.pep	AAFAGVALLMAGGAEEGGEVGWFGCLLVLLAGAGFCAAMRPTQRLIARIGAPAFTSVSIA	180
10	orf62ng		180
		AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGLGC	216
15		AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGLGCGWYAYWLWNKGMSRVPANASGLLI	240
	The complete length	ORF62ng nucleotide sequence <seq 247="" id=""> is:</seq>	
20	51 CGC 101 GCC 151 GGC 201 CAA 251 CCG 301 TTT	TTTTACC AAATCCTTGC CCTGATTATC TGGGGCAGCT CGTTTATTGC CCAAATAT GTCTATGGCG GCATCGATCC CGCATTGATG GTCGGCGTGC CTGCTGAT TGCCGCGCTG CCTGCACTGC CCGCCTGCCG CCGTCATGTC CAAGATTC CGCGTGAGGA ATGGAAGCCG TTGCTGATTG TGTCGTTCGT CCTATGTG CTGACCCTGC TGCTTCAGTT TGTCGGGTTG AAATACACTT CCCGCCAG CGCATCGGTC ATTGTCGGAC TCGAGCCGCT GCTGATGGTG CGTCGGAC ACTTTTTCTT CAACGACAAA GCGCGTGCCT ACCACTGGAT	
25	351 ATG	CGGCGCG GCGCATTTG CCGGTGTCGC GCTGCTGATG GCGGGCGGTG AAGAGGG CGGCGAAGTC GGCTGGTTCG GCTGCTGCT GGTGTTGTTG	
	451 GCG 501 CCG 551 TGA	GGCGCGG GCTTTTGTGC CGCTATGCGT CCGACGCAAA GGCTGATTGC GCATCGGC GCACCGGCAT TCACATCTGT TTCCATTGCC GCCGCATCGT ATGTGCCT GCCGTTTTCG CTTGCTTTGG CGCAAAGTTA TACCGTGGAC GAGCGTCG GGATGGTATT GTCGCTGTTG TATTTGGGTT TGGGGTGCGG	i
30	651 CTG	GTACGCC TATTGGCTGT GGAACAAGGG GATGAGCCGT GTTCCTGCCA GCGTCGGG ACTGTTGATT TCGCTCGAAC CCGTCGTCGG CGTGCTGTTG	
	751 GCG	GTTTTGA TTTTGGGCGA ACATTTATCG CCCGTGTCCG CCTTGGGCGT	ł
		TGTCGTC ATCGCCGCCA CTTTCGCCGC CGGCCGGCTG TCGCGCAGGG	
35	This encodes a prote	ein having amino acid sequence <seq 248="" id="">:</seq>	
	51 GKI 101 <u>FV</u> G	OILALII WGSSFIAAKY VYGGIDPALM VGVRLLIAAL PALPACRRHV PREEWKP LLIVSFVNYV LTLLLQFVGL KYTSAASASV IVGLEPLLMV SHFFFNDK ARAYHWICGA AAFAGVALLM AGGAEEGGEV GWFGCLLVLL GFCAAMR PTORLIARIG APAFTSVSTA AASIMCLPFS LALAOSYTVD	
40	51 GKI 101 <u>FVG</u> 151 <u>AGA</u> 201 WSV	PREEWKP LLIVSFVNYV LTLLLQFVGL KYTSAASASV IVGLEPLLMV SHFFFNDK ARAYHWICGA AAFAGVALLM AGGAEEGGEV GWFGCLLVLL GFCAAMR PTQRLIARIG APAFTSVSIA AASLMCLPFS LALAQSYTVD GMVLSLL YLGLGCGWYA YWLWNKGMSR VPANASGLLI SLEPVVGVLL	
40	51 GKI 101 <u>FVG</u> 151 <u>AGA</u> 201 WSV 251 <u>AVL</u>	PREEWKP LLIVSFVNYV LTLLLQFVGL KYTSAASASV IVGLEPLLMV HFFFNDK ARAYHWICGA AAFAGVALLM AGGAEEGGEV GWFGCLLVLL GFCAAMR PTQRLIARIG APAFTSVSIA AASLMCLPFS LALAQSYTVD GMVLSLL YLGLGCGWYA YWLWNKGMSR VPANASGLLI SLEPVVGVLL LILGEHLS PVSALGVFVV IAATFAAGRL SRRDAQNGNA V*	
40	51 GKI 101 <u>FVG</u> 151 <u>AGA</u> 201 WSV 251 <u>AVL</u>	CHREEWKP LLIVSFVNYV LTLLLQFVGL KYTSAASASV IVGLEPLLMV CHFFFNDK ARAYHWICGA AAFAGVALLM AGGAEEGGEV GWFGCLLVLL CHFFFNDK ARAYHWICGA AAFAGVALLM AGGAEEGGEV GWFGCLLVLL CHGFCAAMR PTQRLIARIG APAFTSVSIA AASLMCLPFS LALAQSYTVD CMVLSLL YLGLGCGWYA YWLWNKGMSR VPANASGLLI SLEPVVGVLL CHGPCLAAMR PTQRLIARIG APAFTSVSIA AASLMCLPFS LALAQSYTVD CMVLSLL YLGLGCGWYA YWLWNKGMSR VPANASGLLI SLEPVVGVLL CHGPCLAAMR PTQRLIARIG AAFAGVALL SRCDAQNGNA V*  52-1 show 97.9% identity in 283 aa overlap:	
	51 GKI 101 <u>FVG</u> 151 <u>AGA</u> 201 WSV 251 <u>AVL</u>	PREEWKP LLIVSFVNYV LTLLLQFVGL KYTSAASASV IVGLEPLLMV HFFFNDK ARAYHWICGA AAFAGVALLM AGGAEEGGEV GWFGCLLVLL HGFCAAMR PTORLIARIG APAFTSVSIA AASLMCLPFS LALAQSYTVD HGMVLSLL YLGLGCGWYA YWLWNKGMSR VPANASGLLI SLEPVVGVLL HILGEHLS PVSALGVFVV IAATFAAGRL SRRDAQNGNA V*  52-1 show 97.9% identity in 283 aa overlap:  10 20 30 40 50  MFYQILALIIWGSSFIAAKYVYGGIDPALMVGVRLLIAALPALPACRRHVGKIPREEWI	50 (P
40	51 GKI 101 FVG 151 AGA 201 WSV 251 AVL ORF62ng and ORF6	PREEWKP LLIVSFVNYV LTLLLQFVGL KYTSAASASV IVGLEPLLMV SHFFFNDK ARAYHWICGA AAFAGVALLM AGGAEEGGEV GWFGCLLVLL GFCAAMR PTQRIARIG APAFTSVSTA AASLMCLPFS LALAQSYTVD GMVLSLL YLGLGCGWYA YWLWNKGMSR VPANASGLLI SLEPVVGVLL JILGEHLS PVSALGVFVV IAATFAAGRL SRRDAQNGNA V*  52-1 show 97.9% identity in 283 aa overlap:  10 20 30 40 50  MFYQILALIIWGSSFIAAKYVYGGIDPALMVGVRLLIAALPALPACRRHVGKIPREEWI	KP
	51 GKI 101 EVG 151 AGA 201 WSV 251 AVL ORF62ng and ORF6	PREEWKP LLIVSFVNYV LTLLLQFVGL KYTSAASASV IVGLEPLLMV HFFFNDK ARAYHWICGA AAFAGVALLM AGGAEEGGEV GWFGCLLVLL HGFCAAMR PTQRIIARIG APAFTSVSIA AASLMCLPFS LALAQSYTVD HGMVLSLL YLGLGCGWYA YWLWNKGMSR VPANASGLLI SLEPVVGVLL HILGEHLS PVSALGVFVV HAATFAAGRL SRRDAQNGNA V*  52-1 show 97.9% identity in 283 aa overlap:  10 20 30 40 50  MFYQILALIIWGSSFIAAKYVYGGIDPALMVGVRLLIAALPALPACRRHVGKIPREEWI  HILIHIIHIHIHIHIHIHIHIHIHIHIHIHIHIHIHI	KP
45	51 GKI 101 <u>FVG</u> 151 <u>AGA</u> 201 WSV 251 <u>AVL</u> ORF62ng and ORF6 orf62ng.pep	PREEWKP LLIVSFVNYV LTLLLQFVGL KYTSAASASV IVGLEPLLMV CHFFFNDK ARAYHWICGA AAFAGVALLM AGGAEEGGEV GWFGCLLVLL GGFCAAMR PTQRLIARIG APAFTSVSIA AASLMCLPFS LALAQSYTVD GMVLSLL YLGLGCGWYA YWLWNKGMSR VPANASGLLI SLEPVVGVLL LILGEHLS PVSALGVFVV IAATFAAGRL SRRDAQNGNA V*  52-1 show 97.9% identity in 283 aa overlap:  10 20 30 40 50  MFYQILALIIWGSSFIAAKYVYGGIDPALMVGVRLLIAALPALPACRRHVGKIPREEWI 11 11 11 11 11 11 11 11 11 11 11 11 11	CP II CP 50
	51 GKI 101 FVG 151 AGA 201 WSV 251 AVL ORF62ng and ORF6 orf62ng.pep orf62-1	PREEWKP LLIVSFVNYV LTLLLQFVGL KYTSAASASV IVGLEPLLMV CHFFFNDK ARAYHWICGA AAFAGVALLM AGGAEEGGEV GWFGCLLVLL GGFCAAMR PTQRLIARIG APAFTSVSIA AASLMCLPFS LALAQSYTVD CMVLSLL YLGLGCGWYA YWLWNKGMSR VPANASGLLI SLEPVVGVLL LILGEHLS PVSALGVFVV IAATFAAGRL SRRDAQNGNA V*  62-1 show 97.9% identity in 283 aa overlap:  10 20 30 40 50  MFYQILALIIWGSSFIAAKYVYGGIDPALMVGVRLLIAALPALPACRRHVGKIPREEWI 11 11 11 11 11 11 11 11 11 11 11 11 11	KP KP 50 20 5A
45	51 GKI 101 <u>FVG</u> 151 <u>AGA</u> 201 WSV 251 <u>AVL</u> ORF62ng and ORF6 orf62ng.pep	PREEWKP LLIVSFVNYV LTLLLQFVGL KYTSAASASV IVGLEPLLMV SHFFFNDK ARAYHWICGA AAFAGVALLM AGGAEEGGEV GWFGCLLVLL GGFCAAMR PTQRLIARIG APAFTSVSIA AASLMCLPFS LALAQSYTVD GMVLSLL YLGLGCGWYA YWLWNKGMSR VPANASGLLI SLEPVVGVLL LILGEHLS PVSALGVFVV IAATFAAGRL SRRDAQNGNA V*  52-1 show 97.9% identity in 283 aa overlap:  10 20 30 40 50  MFYQILALIIWGSSFIAAKYVYGGIDPALMVGVRLLIAALPALPACRRHVGKIPREEWI 11 11 11 11 11 11 11 11 11 11 11 11 11	KP KP 50 20 5A
45	51 GKI 101 FVG 151 AGA 201 WSV 251 AVL  ORF62ng and ORF6  orf62ng.pep orf62-1  orf62ng.pep	PREEWKP LLIVSFVNYV LTLLLQFVGL KYTSAASASV IVGLEPLLMV SHFFFNDK ARAYHWICGA AAFAGVALLM AGGAEEGGEV GWFGCLLVLL GFCAAMR PTQRIARIG APAFTSVSIA AASLMCLPFS LALAQSYTVD GMVLSLL YLGLGCGWYA YWLWNKGMSR VPANASGLLI SLEPVVGVLL LILGEHLS PVSALGVFVV IAATFAAGRL SRRDAQNGNA V*  52-1 show 97.9% identity in 283 aa overlap:  10 20 30 40 50  MFYQILALIIWGSSFIAAKYVYGGIDPALMVGVRLLIAALPALPACRRHVGKIPREEWI  MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLLIAALPALPACRRHVGKIPREEWI 10 20 30 40 50  70 80 90 100 110 1:  LLIVSFVNYVLTLLLQFVGLKYTSAASASVIVGLEPLLMVFVGHFFFNDKARAYHWICG 111111111111111111111111111111111111	CP CP CP 50 20 GA 11 GA 20
<b>45</b> <b>50</b>	51 GKI 101 FVG 151 AGA 201 WSV 251 AVL ORF62ng and ORF6 orf62ng.pep orf62-1	PREEWKP LLIVSFVNYV LTLLLQFVGL KYTSAASASV IVGLEPLLMV SHFFFNDK ARAYHWICGA AAFAGVALLM AGGAEEGGEV GWFGCLLVLL GFCAAMR PTQRIARIG APAFTSVSIA AASLMCLPFS LALAQSYTVD GMVLSLL YLGLGCGWYA YWLWNKGMSR VPANASGLLI SLEPVVGVLL LILGEHLS PVSALGVFVV IAATFAAGRL SRRDAQNGNA V*  52-1 show 97.9% identity in 283 aa overlap:  10 20 30 40 50  MFYQILALIIWGSSFIAAKYVYGGIDPALMVGVRLLIAALPALPACRRHVGKIPREEWI	(P         (P   50   50   1     53   1     53   60   61   61   63   63   63   63   63   63   63   63
<b>45</b> <b>50</b>	51 GKI 101 FVG 151 AGA 201 WSV 251 AVL  ORF62ng and ORF6  orf62ng.pep orf62-1  orf62ng.pep	PREEWKP LLIVSFVNYV LTLLLQFVGL KYTSAASASV IVGLEPLLMV HFFFNDK ARAYHWICGA AAFAGVALLM AGGAEEGGEV GWFGCLLVLL HGFCAAMR PTQRIARIG APAFTSVSIA AASLMCLPFS LALAQSYTVD HGWVLSLL YLGLGCGWYA YWLWNKGMSR VPANASGLLI SLEPVVGVLL HILGEHLS PVSALGVFVV IAATFAAGRL SRRDAQNGNA V*  52-1 show 97.9% identity in 283 aa overlap:  10 20 30 40 50  MFYQILALIIWGSSFIAAKYVYGGIDPALMVGVRLLIAALPALPACRRHVGKIPREEWI MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLLIAALPALPACRRHVGKIPREEWI 10 20 30 40 50  70 80 90 100 110 1:  LLIVSFVNYVLTLLLQFVGLKYTSAASASVIVGLEPLLMVFVGHFFFNDKARAYHWICG 11 130 140 150 160 170 1:  130 140 150 160 170 1:  131 140 150 160 170 1:  AAFAGVALLMAGGAEEGGEVGWFGCLLVLLAGAGFCAAMRPTQRLIARIGAPAFTSVS:  11 11 11 11 11 11 11 11 11 11 11 11 11	KP KP 500 200 3A 111 5A 200 31A 111
<b>45</b> <b>50</b>	51 GKI 101 FVG 151 AGA 201 WSV 251 AVL  ORF62ng and ORF6  orf62ng.pep orf62-1  orf62ng.pep orf62-1	PREEWKP LLIVSFVNYV LTLLLQFVGL KYTSAASASV IVGLEPLLMV HFFFNDK ARAYHWICGA AAFAGVALLM AGGAEEGGEV GWFGCLLVLL HGFFNDK ARAYHWICGA AAFAGVALLM AGGAEEGGEV GWFGCLLVLL HGFCAAMR PTQRLIARIG APAFTSVSIA AASLMCLPFS LALAQSYTVD HGMVLSLL YLGLGCGWYA YWLWNKGMSR VPANASGLLI SLEPVVGVLL HILGEHLS PVSALGVFVV IAATFAAGRL SRRDAQNGNA V*  52-1 show 97.9% identity in 283 aa overlap:  10	(P (CP (CP (CP (CP (CP (CP (CP (CP (CP (
45 50 55	51 GKI 101 FVG 151 AGA 201 WSV 251 AVL  ORF62ng and ORF6  orf62ng.pep orf62-1  orf62ng.pep orf62-1	PREEWKP LLIVSFVNYV LTLLLQFVGL KYTSAASASV IVGLEPLLMV HIFFFNDK ARAYHWICGA AAFAGVALLM AGGAEEGGEV GWFGCLLVLL GGFCAAMR PTQRLIARIG APAFTSVSIA AASLMCLPFS LALAQSYTVD GMVLSLL YLGLGCGWYA YWLWNKGMSR VPANASGLLI SLEPVVGVLL LILGEHLS PVSALGVFVV IAATFAAGRL SRRDAQNGNA V*  52-1 show 97.9% identity in 283 aa overlap:  10 20 30 40 50  MFYQILALIIWGSSFIAAKYVYGGIDPALMVGVRLLIAALPALPACRRHVGKIPREEWI 11 11 11 11 11 11 11 11 11 11 11 11 11	(P (CP (CP (CP (CP (CP (CP (CP (CP (CP (
45 50 55	51 GKI 101 FVG 151 AGA 201 WSV 251 AVL  ORF62ng and ORF6  orf62ng.pep orf62-1  orf62ng.pep orf62-1	PREEWKP LLIVSFVNYV LTLLLQFVGL KYTSAASASV IVGLEPLIMV BHFFFNDK ARAYHWICGA AAFAGVALIM AGGAEEGGEV GWFGCLLVLL GFCAAMR PTQRLIARIG APAFTSVSIA AASLMCLPFS LALAQSYTVD IGMVLSLL YLGLGCGWYA YWLWNKGMSR VPANASGLLI SLEPVVGVLL JILGEHLS PVSALGVFVV IAATFAAGRL SRRDAQNGNA V*  52-1 show 97.9% identity in 283 aa overlap:  10 20 30 40 50  MFYQILALIIWGSSFIAAKYVYGGIDPALMVGVRLLIAALPALPACRRHVGKIPREEWI 10 20 30 40 50  MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLLIAALPALPACRRHVGKIPREEWI 10 20 30 40 50  70 80 90 100 110 1:  LLIVSFVNYVLTLLLQFVGLKYTSAASASVIVGLEPLLMVFVGHFFFNDKARAYHWICG 11 11 11 11 11 11 11 11 11 11 11 11 11	KP KP 50 20 3A 11 30 11 11 11

WO 99/24578 PCT/IB98/01665

-180-

```
250
                                  260
                                          270
                                                  280
                                                          290
                    SLEPVVGVLLAVLILGEHLSPVSALGVFVVIAATFAAGRLSRRDAQNGNAVX
        orf62ng.pep
                    SLEPVVGVLLAVLILGEHLSPVSALGVFVVIAATLVAGRLSHQKX
5
        orf62-1
                         250
                                  260
                                          270
```

Furthermore, ORF62ng shows significant homology to a hypothetical *H.influenzae* protein:

```
sp|Q57147|Y976_HAEIN HYPOTHETICAL PROTEIN HI0976 >gi|1074589|pir||B64163
          hypothetical protein HI0976 - Haemophilus influenzae (strain Rd KW20)
10
          >gi|1574004 (U32778) hypothetical [Haemophilus influenzae] Length = 128
           Score = 106 bits (262), Expect = 2e-22
           Identities = 56/114 (49%), Positives = 68/114 (59%)
                     MFYQILALIIWGSSFIAAKYVYGGIDPALMVGVRXXXXXXXXXXXCRRHVGKIPREEWKP 60
           Query: 1
15
                      M YQILAL+IW SS I K Y +DP L+V VR
                                                                    R
                                                                        KI +
           Sbict: 1
                     MLYQILALLIWSSSLIVGKLTYSMMDPVLVVQVRLIIAMIIVMPLFLRRWKKIDKPMRKO 60
           Query: 61
                     LLIVSFVNYVLTLLLQFVGLKYTSAASASVIVGLEPLLMVFVGHFFFNDKARAY 114
                                  LLQF+GLKYTSA+SA ++GLEPLL+VFVGHFFF K
                      L ++F NY
20
           Sbjct: 61 LWWLAFFNYTAVFLLQFIGLKYTSASSAVTMIGLEPLLVVFVGHFFFKTKONGF 114
```

Based on this analysis, including the homology with the transmembrane protein of H.influenzae and the putative leader sequecne and several transmembrane domains in the gonococcal protein, it is predicted that these proteins from N.meningitidis and N.gonorrhoeae, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 30

25

The following partial DNA sequence was identified in N.meningitidis <SEO ID 249>:

```
ATGCGCCGTT TTCTACCGAT CGCAGCCATA TGCGCmGwms TCCTGkkGTA
                51
                    SGGACTGACG GCGGCAACCG GCAGCACCAG TTCGCTGGCG GATTATTTCT
30
                    GGTGGATTGT TGCGTTCAGC GCAATGCTGC TGCTGGTGTT GTCCGCCGTT
               101
               151
                    TTGGCACGTT ATGTCATATT GCTGTTGAAA GACAGGCGCG ACGGCGTATT
               201
                    CGGTTCGCtA srTyGCCAAA gsGCCTgkks TGGG.ATGTT TACGCTGGTT
                    GCCGKACTGC CCGGCGTGTT TCTGTTCGGC TTTCCCGCAC AGTTCATCAA
               251
                    CGGCACGATT AATTCGTGGT TCGGCAACGA TACCCACGAG GCGCTTGAAC
               301
35
                    GCAGCCTCAA TTTGAGCAAG TCCGCATTGA ATTTGGCGGC AGACAACGCC
               351
                    CTCGGCAACG CCGTCCCCGT GCAGATAGAC CTCATCGGCG CGGCTTCCCT
               401
               451
                    GCCCGGGGAT ATGGGCAGGG TGCTGGAACA TTACGCCGGC AGCGGTTTTG
               501
                    CCCAGCTTGC CCTGTACAAy ksCGCAAGCG GCAAAATCGA AAAAAGCATC
               551 AACCCGCACA AGCTCGATCA GCCGTTTCCA GGTAAGGCGC GTTGGGAaAa
40
                    AATCCAACGG GCGGGTTCGG TCAGGGATTT GGAAAGCATA GGCGGCGTAT
               601
               651
                    TGTaCGCGCA GGGCTGGCTG TCGGCGGGTA CGCACWACGG GCGCGATTAC
               701
                    GCCTTGTTTT TCCGTCAGCC GGTTCCCAAA GGCGTGGCAG AGGATGCCGT
               751
                    AAGGTTTGCA GACCTTTTTC CTGGCAACCC TGCTGATTGC CTCGCTGCTG
               801
45
               851
                    TCGATTTTC TTGCACTGGT CATGGCACTG TATTTCGCCC GCCGTTTCGT
                    CGAACCCGTC CTATCGCTTG CCGAGGGGGC GAAGGCGGTG GCGCAAGGCG
               901
               951
                    ATTTCAGCCA GACGCCCCC GTGTTGCGCA ACGACGAGTT CGGACGCTTG
                    ACCArGTTGT TCAACCACAT GACCGAGCAG CTTTCCATCG CCAAAGATGC
              1001
              1051
                    AGACGAGCGC AACCGCCGGC GCGAGGAAGC CGCCAGGCAT TATCTTGAAT
50
                    GCGTGTTGGA GGGGCTGACC ACGGGCGTGG TGGTGTTTGA CGAACAAGGC
              1101
              1151
                    TGTCTGAAAA CCTTCAACAA AGCGGCGGGT ACC..
```

This corresponds to the amino acid sequence <SEQ ID 250; ORF64>:

```
MRRFLPIAAI CAXXLXXGLT AATGSTSSLA DYFWWIVAFS AMLLLVLSAV
                 1
                51
                    LARYVILLLK DRRDGVFGSX XAKXPXXXMF TLVAXLPGVF LFGFPAQFIN
55
                101
                    GTINSWFGND THEALERSLN LSKSALNLAA DNALGNAVPV QIDLIGAASL
                     PGDMGRVLEH YAGSGFAQLA LYNXASGKIE KSINPHKLDQ PFPGKARWEK
               151
                201
                     IQRAGSVRDL ESIGGVLYAQ GWLSAGTHXG RDYALFFRQP VPKGVAEDAV
                    LIEKARAKYA ELSYSKKGLO TFFLATLLIA SLLSIFLALV MALYFARRFV
```

301 EPVLSLAEGA KAVAQGDFSQ TRPVLRNDEF GRLTXLFNHM TEQLSIAKDA 351 DERNRREEA ARHYLECVLE GLTTGVVVFD EQGCLKTFNK AAGT..

#### Further work revealed the complete nucleotide sequence <SEQ ID 251>:

```
1 ATGCGCCGTT TTCTACCGAT CGCAGCCATA TGCGCCGTCG TCCTGTTGTA
                    CGGACTGACG GCGGCAACCG GCAGCACCAG TTCGCTGGCG GATTATTTCT
5
                    GGTGGATTGT TGCGTTCAGC GCAATGCTGC TGCTGGTGTT GTCCGCCGTT
               101
                    TTGGCACGTT ATGTCATATT GCTGTTGAAA GACAGGCGCG ACGGCGTATT
               151
                    CGGTTCGCAG ATTGCCAAAC GCCTTTCTGG GATGTTTACG CTGGTTGCCG
               201
                    TACTGCCCGG CGTGTTTCTG TTCGGCGTTT CCGCACAGTT CATCAACGGC
               251
                    ACGATTAATT CGTGGTTCGG CAACGATACC CACGAGGCGC TTGAACGCAG
10
               301
                    CCTCAATTTG AGCAAGTCCG CATTGAATTT GGCGGCAGAC AACGCCCTCG
               351
                    GCAACGCCGT CCCCGTGCAG ATAGACCTCA TCGGCGCGGC TTCCCTGCCC
               401
                    GGGGATATGG GCAGGGTGCT GGAACATTAC GCCGGCAGCG GTTTTGCCCA
               451
                    GCTTGCCCTG TACAATGCCG CAAGCGGCAA AATCGAAAAA AGCATCAACC
               501
15
                    CGCACAAGCT CGATCAGCCG TTTCCAGGTA AGGCGCGTTG GGAAAAAATC
               551
                    CAACGGGCGG GTTCGGTCAG GGATTTGGAA AGCATAGGCG GCGTATTGTA
               601
               651
                    CGCGCAGGGC TGGCTGTCGG CGGGTACGCA CAACGGGCGC GATTACGCCT
                    TGTTTTTCCG TCAGCCGGTT CCCAAAGGCG TGGCAGAGGA TGCCGTCTTA
               701
                    ATCGAAAAGG CAAGGGCGAA ATATGCTGAG TTGAGTTACA GCAAAAAAGG
               751
                    TTTGCAGACC TTTTTCCTGG CAACCCTGCT GATTGCCTCG CTGCTGTCGA
20
               801
                    TTTTTCTTGC ACTGGTCATG GCACTGTATT TCGCCCGCCG TTTCGTCGAA
               851
                    CCCGTCCTAT CGCTTGCCGA GGGGGGGAAG GCGGTGGCGC AAGGCGATTT
               901
               951
                    CAGCCAGACG CGCCCCGTGT TGCGCAACGA CGAGTTCGGA CGCTTGACCA
                    AGTTGTTCAA CCACATGACC GAGCAGCTTT CCATCGCCAA AGAAGCAGAC
              1001
25
              1051
                    GAGCGCAACC GCCGGCGCA GGAAGCCGCC AGGCATTATC TTGAATGCGT
                    GTTGGAGGGG CTGACCACGG GCGTGGTGGT GTTTGACGAA CAAGGCTGTC
              1101
              1151
                    TGAAAACCTT CAACAAAGCG GCGGAACAGA TTTTGGGGAT GCCGCTTACC
                    CCCTGTGGG GCAGCAGCCG GCACGGTTGG CACGGCGTTT CGGCGCAGCA
              1201
                    1251
30
                    ACAAACCGGT CCATGTGAAA TATGCCGCGC CGGACGATGC CAAAATCCTG
               1301
                    CTGGGCAAGG CAACCGTCCT GCCCGAAGAC AACGGCAACG GCGTGGTAAT
               1351
                     GGTGATTGAC GACATCACCG TTTTGATACA CGCGCAAAAA GAAGCCGCGT
               1401
                     GGGGCGAAGT GGCGAAGCGG CTGGCACACG AAATCCGCAA TCCGCTCACG
               1451
                     CCCATCCAGC TTTCCGCCGA ACGGCTGGCG TGGAAATTGG GCGGGAAGCT
               1501
                     GGATGAGCAG GATGCGCAAA TCCTGACGCG TTCGACCGAC ACCATCGTCA
35
               1551
                     AACAGGTGGC GGCATTGAAG GAAATGGTCG AAGCATTCCG CAATTATGCG
               1601
                     CGTTCCCCTT CGCTCAAATT GGAAAATCAG GATTTGAACG CCTTAATCGG
               1651
                     CGATGTGTTG GCATTGTATG AAGCCGGTCC GTGCCGGTTT GCGGCGGAGC
               1701
               1751
                     TTGCCGGCGA ACCGCTGACG GTGGCGGCGG ATACGACCGC CATGCGGCAG
                     GTGCTGCACA ATATTTCAA AAATGCCGCC GAAGCGGCGG AAGAAGCCGA
40
               1801
                     TGTGCCCGAA GTCAGGGTAA AATCGGAAAC AGGGCAGGAC GGTCGGATTG
               1851
                     TCCTGACGGT TTGCGACAAC GGCAAAGGGT TCGGCAGGGA AATGCTGCAC
               1901
                     AACGCCTTCG AGCCGTATGT AACGGACAAA CCGGCGGGAA CGGGATTGGG
               1951
                     TCTGCCTGTG GTGAAAAAA TCATTGAAGA ACACGGCGGC CGCATCAGCC
               2001
 45
                     TGAGCAATCA GGATGCGGGT GGCGCGTGTG TCAGAATCAT CTTGCCAAAA
               2051
                     ACGGTAAAAA CTTATGCGTA G
               2101
```

#### This corresponds to the amino acid sequence <SEQ ID 252; ORF64-1>:

	1 ·	MRRFLPIAAI	CAVVLLYGLT	AATGSTSSLA	DYFWWIVAFS	AMLLLVLSAV
	51	LARYVILLLK	DRRDGVFGSQ	IAKRLSGMFT	LVAVLPGVFL	FGVSAQFING
50	101				NALGNAVPVQ	
• -	151	GDMGRVLEHY	AGSGFAQLAL	YNAASGKIEK	SINPHKLDQP	FPGKARWEKI
	201	QRAGSVRDLE	SIGGVLYAQG	WLSAGTHNGR	DYALFFRQPV	PKGVAEDAVL
	251	IEKARAKYAE	LSYSKKGLQT	FFLATLLIAS	LLSIFLALVM	ALYFARREVE
	301	PVLSLAEGAK	AVAQGDFSQT	RPVLRNDEFG	RLTKLFNHMT	EQLSIAKEAD
55	351	ERNRRREEAA	RHYLECVLEG	LTTGVVVFDE	QGCLKTFNKA	AEQILGMPLT
	401	PLWGSSRHGW	HGVSAQQSLL	<b>AEVFAAIGAA</b>	AGTDKPVHVK	YAAPDDAKIL
	451	LGKATVLPED	NGNGVVMVID	DITVLIHAQK	EAAWGEVAKR	LAHEIRNPLT
	501	PIOLSAERLA	WKLGGKLDEQ	DAQILTRSTD	TIVKQVAALK	<b>EMVEAFRNYA</b>
	551	RSPSLKLENO	DLNALIGDVL	ALYEAGPCRF	AAELAGEPLT	VAADTTAMRQ
60	601	VLHNIFKNAA	EAAEEADVPE	VRVKSETGQD	GRIVLTVCDN	GKGFGREMLH
00	651	NAFEPYVTDK	PAGTGLGLPV	VKKIIEEHGG	RISLSNQDAG	GACVRIILPK
	701	TVKTYA*			_	

Computer analysis of this amino acid sequence gave the following results:

# Homology with a predicted ORF from N. meningitidis (strain A)

ORF64 shows 92.6% identity over a 392aa overlap with an ORF (ORF64a) from strain A of N. meningitidis:

	meningitidis:		
		10 20 30 40 50	60
_		10 20 30 40 MRRFLPIAAICAXXLXXGLTAATGSTSSLADYFWWIVAFSAMLLLVLSAVLARYVILL	<u>L</u> K
5	orf64.pep	- <del>                                     </del>	, ,
	orf64a	MPDFI PTAATCAVVILLYGLTAATGSTSSLADYFWWIVAFSAMLLLVLSAVLARYVILL	TK
	OIID4a	10 20 30 40 50	60
			20
10		70 80 100 410 -	.20
10	orf64.pep	DRRDGVFGSXXAKXPXXXMFTLVAXLPGVFLFGFPAQFINGTINSWFGNDTHEALERS	LI
	• • •		II.N
	orf64a	DRRDGVFGSOIAKR-LSGMFTLVAVLPGVFLFGVSAQFINGTINSWFGNDTHEALERS	
		70 80 90 100 110	
15			180
	ee4	TOWARD NEAD NATIONAL CHANDOOT DITGAASLPGDMGRVLEHYAGSGFAQLALYNXASGF	<b>KIE</b>
	orf64.pep		111
	orf64a	I.SKSDI.NI.AADNALGNAIPVOIDXIGAASLPXDMGRVLEHIAGSGEAQLALINAASG	KIE
20	OFIO4a	120 130 140 150 160 170	
20			040
			240
	orf64.pep	KSINPHKLDQPFPGKARWEKIQRAGSVRDLESIGGVLYAQGWLSAGTHXGRDYALFF	KQP IIII
		ASINFHALDOFF FOR AN	ROP
25	orf64a	KSINPHKLDOPFPGKARWEKIQQAGSVRDXESIGGVLYAXGWLSAXTHNGRDYALFF	
		180 190 200 210 220 230	
		250 260 270 280 290	300
		250 260 270 280 290 VPKGVAEDAVLIEKARAKYAELSYSKKGLQTFFLATLLIASLLSIFLALVMALYFAR	RFV
20	orf64.pep		111
30	orf64a	VPKCVAEDAVI.IEKARAXXXXLSYSKKGLQTFFLATLLIASLLSIFLALVMALYFAR	RFV
	OLIO4a	240 250 260 270 280 290	
		000	360
		310 320 330 340 350	
35	orf64.pep	EPVLSLAEGAKAVAQGDFSQTRPVLRNDEFGRLTXLFNHMTEQLSIAKDADERNRRF	III
-	• •		₹EEA
	orf64a	EPVLSLAEGAKAVAQGDFSQTRPVLRNDEFGRLTKLFNHMTEQLSIAKEADERNRRF	
		300 310 320 330 340 350	
40		370 380 390	
40		ARHYLECVLEGLTTGVVVFDEQGCLKTFNKAAGT	
	orf64.pep		
	orf64a	ARHYLECVLEGLTTGVVVFDEQGCLKTFNKAAEQILGMPLTPLWGSSRHGWHGVSA	<b>JOST</b>
	022 4 4 -	360 370 380 390 400 410	
45		TO THE PROPERTY OF THE PROPERT	OAHT
• •	orf64a	LAEVFAAIGAAAGTDKPVHVKYAAPDDAKILLGKATVLPEDNXNGVVMVIDDITVL	
		420 430 440 430	
	The complete le	ngth ORF64a nucleotide sequence <seq 253="" id=""> is:</seq>	
	The complete ici	ngm Old Old named to I amend	
	1	ATGCGCCGTT TTCTACCGAT CGCAGCCATA TGCGCCGTCG TCCTGTTGTA	
50		CCCACTCACC CCCCCAACCC CCAGCACCAG TTCGCTGGCG GATTATTTCT	
30	101	CONCOMMON TECCHTEAGE GEAATGETEC TGETGGTGT GTCCGCCGII	
	151	MMCCCACCMT ATCTCATATT CCTGTTGAAA GACAGGCGCG ACGGCGTATT	
	201	CCCTTCCCAC ATTCCCAAAC GCCTTTCCGG GATGTTTACG CTGGTTGCCG	
	251	TACTGCCCGG CGTGTTTCTG TTCGGCGTTT CCGCACAGTT TATCAACGGC	
55	301	ACGATTAATT CGTGGTTCGG CAACGATACC CACGAGGCGC TTGAACGCAG	
-	351	CCTCAATTTG AGCAAGTCCG CATTGAATCT GGCGGCAGAC AACGCCCTTG	
	401	GCAACGCCAT CCCCGTGCAG ATAGACNTCA TCGGCGCGGC TTCCCTGCCC NGGGATATGG GCAGGGTGCT GGAACATTAC GCCGGCAGCG GTTTTGCCCA	
	451	NGGGATATGG GCAGGGTGCT GGAACATTAC GCCGGCAGA AGCATCAACC GCTTGCCCTG TACAATGCCG CAAGCGGCAA AATCGAAAAA AGCATCAACC	
	501	CGCACAAGCT CGATCAGCCG TTTCCAGGTA AGGCGCGTTG GGAAAAAATC	
60		CARCAGGGG GTTCGGTCAG GGATNNGGAA AGCATAGGCG GCGTATTGIA	
	601 651	CCCCANCEC TECTTETEE CAGNNACGEA CAACGGGCGC GATTACGCCT	
	701	MCMMMMTCCG TCAGCCGGTT CCCAAAGGCG TGGCAGAGGA TGCCGTCTA	
	701 751	AMOCANANCE CARGEGEENA ANANNNTNAG TTGAGTTACA GCAAAAAAGG	
65		MARICACACC TETTECCTNG CAACCCTGCT GATTGCCTCN CTGCTGTCGA	
0.	851		
	331		

			•				
	901	CCCGTCCTAT	CGCTTGCCGA	GGGGGCGAAG	GCGGTGGCGC	AAGGCGATTT	
	951	CAGCCAGACG	CGCCCCGTGT	TGCGCAACGA	CGAGTTCGGA	CGCTTGACCA	
	1001				CCATCGCCAA		
_	1051				AGACATTATC		
5	1101				GTTTGACGAA		
	1151	TGAAAACCTT	CAACAAAGCG	CCACCCTTCC	TTTTGGGGAT	CGGCGCAGCA	
	1201	CCCCTGTGGG	CCCCAACTCT	TTCCCCCCAT	CGGCGCGCG	CCACCTACCA	
	1251 1301					CAAAATCCTG	
10	1301	CTCCCCAACC	CANGIGIAAA	GCCCGAAGAC	AACNGCAACG	GCGTGGTAAT	
10	1401	CTGGGCAAGG	GACATCACCG	TTTTGATACA	CGCGCAAAAA	GAAGCCGCGT	
	1451					TCCGCTCACG	
	1501					GCGGGAAGCT	
	1551					ACCATCATCA	
15	1601	AACAAGTGGC	GGCATTAAAA	GAAATGGTCG	AGGCATTCCG	CAATTACNCG	
	1651					CCTTAATCGG	
	1701					GCGGCGGAAC	
	1751	TTGCCGGCGA	ACCGCTGATG	ATGGCGGCGG	ATACGACCGC	CATGCGGCAG	
	1801					AAGAAGCCGA	
20	1851					GGACGGATTG	
	1901	TCCTGACAGT	TTGCGACAAC	A A CCCA CA A A	CCCCCTCCAA	AATGCTGCAC CGGGATTGNG	
	1951	AATGCCTTCG	AGCCGTATGI	TO T	ACACCCCCCC	CNCATCAGCC	•
	2001 2051	MCACCA ATCA	CCATCCGGGC	CCCCCCTNTC	TCACGGCGGC	CTTGCCAAAA	
25	2101		CTTATGCGTA		TCAGARICAI	CIIGCOMM	
23							
	This encodes a p	orotein havin	ig amino acid	d sequence <	SEQ ID 254	<b>!&gt;:</b>	!
			J	•	•		,
	1	MRRFLPIAAI	CAVVLLYGLT	AATGSTSSLA	DYFWWIVAFS	AMLLLVLSAV	
	51	LARYVILLLK	DRRDGVFGSQ	IAKRLSGMFI	LVAVLPGVFI	FGVSAQFING	
	101	TINSWFGNDT	HEALERSLNL	SKSALNLAAD	NALGNAIPVO	IDXIGAASLP	
30	151					FPGKARWEKI	
•	201					PKGVAEDAVL	,
	251					ALYFARREVE	
	301	PVLSLAEGAK	AVAQGDFSQI	RPVLRNDEF	RLTKLFNHMI	EQLSIAKEAD	
25	351	EKNRKKEEAA	KHIPECAPEG	LITGVVVEDE	O ACCIVILING	AEQILGMPLT YAAPDDAKIL	
35	401 451	PLWGSSKRGW	NANGWWMILL HRASWÖÖSPT	AEVEAALGAA	C EDDMCEADAL	LAHEIRNPLT	
	501	DIGITAL VIELE	WKIGGKIDEX	DAOTLTEST	TIIKOVAALI	EMVEAFRNYX	
	551	RSPSXOLENC	DLNALIGDVI	ALYEAGPCRE	AAELAGEPLN	MAADTTAMRQ	
	601	VLHNIFKNA	EAAEEADVPE	VRVKSEAGQI	GRIVLTVCD	GKGFGREMLH	
40	651					GAXVRIILPK	
	701	TVETYA*					
	07764 101	DDC4 1 1	. 0.0 (0/ : 4		1		
	ORF64a and Ol	RF64-1 shov	v 96.6% ider	itity in 706 a	a overlap:		
			10		30 40		60
	orf64a.pe	MRRFLPI				SAMLLLVLSAVL/	
45		11(11)					
	orf64-1	MRRFLP.			30 41		60
			10	20 .		, , ,	
			70 ·	80	90 10	110	120
50	orf64a.pe	en DRRDGVI				GTINSWFGNDTH	
50	OZZVIGIP	111111		11111111111	пинийн		(11111111
	orf64-1	DRRDGV	FGSQIAKRLSG	1FTLVAVLPGV	FLFGVSAQFIN	GTINSWFGNDTH	EALERSLNL
		-	70	80	90 10	0 110	120
55					50 16		180
	orf64a.pe	ep SKSALN	LAADNALGNAI	PVQIDXIGAAS:	LPXDMGRVLEH	YAGSGFAQLALY	NAASGKIEK
						пинини	
	orf64-1	SKSALN				YAGSGFAQLALY	
			130	140 1	50 16	0 170	180
60	•			000 0	10 00	0 20	240
		on CTNDIII			10 22 XESTGGVLVAX	0 230 GWLSAXTHNGRD	
	orf64a.p	ећ этикик	TILLIIIIII. TUUEEEGKAKW	•       			
	orf64-1	SINDHR	LDOPFPGKARW	EKIORAGSVRD	LESIGGVLYAO	GWLSAGTHNGRD	YALFFROPV
65	OITOA	JIHTHK			10 22		240
0,5			- <del>-</del> -	_		· •	•
		•	250	260 2	70 28	0 290	300
•							

_	orf64a.pep	111111	111111111	111111	GLQTFFLATLI            GLQTFFLATLI 270	111111111	11111111	1111
5	orf64a.pep	111111	11111111111	1111111111	330 DEFGRLTKLFN			1111
10	orf64-1	PVLSLA	EGAKAVAQGD 310	FSQTRPVLRN 320	DEFGRLTKLF1 330	HMTEQLSIAN 340	KEADERNRRR 350	360 420
	orf64a.pep	111111	THEFT	HILLIAM	390 FNKAAEQILGI	1111111111	111111111	QSLL
15	orf64-1	RHYLEC	VLEGLTTGVV 370	VFDEQGCLKT 380	FNKAAEQILGI 390	400	410	420
20	orf64a.pep	111111	1111111111	11111111111	450 AKILLGKATV           AKILLGKATV	!!!!! !!!!	11111111	
	orf64-1	ALVIA	430	440	450	460	470 530	480 540
25	orf64a.pep	111111	11111111111		S10 AERLAWKLGGK           AERLAWKLGGK 510	LDEXDAQILT	RSTDTIIKQ'	VAALK
30	orf64a.pep	11111	: FRNYARSPSLI	:          KLENQDLNAL:	570 IGDVLALYEAG             IGDVLALYEAG		111:111	1111
35	orf64a.pep	<b>VLHNI</b>	550 610 FKNAAEAAEE	560 620 ADVPEVRVKSI	570 630 EAGQDGRIVLI	580 640 VCDNGKGFGF	650 REMLHNAFEP	660 YVTDK
40	orf64-1	11111	1111111111	11111111	:         ETGQDGRIVLT 630	1111111111	:	11111
	orf64a.pep	11111	1 11111111	11111 1111	690 NQDAGGAXVRI		11	
45	orf64-1	PAGTG	LGLPVVKKII 670	EEHGGRISLS 680	NODAGGACVR: 690	700	ΑX	

## Homology with a predicted ORF from N.gonorrhoeae

ORF64 shows 86.6% identity over a 387aa overlap with a predicted ORF (ORF64.ng) from N.

## 50 gonorrhoeae:

	orf64.pep	MRRFLPIAAICAXXLXXGLTAATGSTSSLADYFWWIVAFSAMLLLVLSAVLARYVILLLK	60
	orf64ng	MRRFLPIAAICAVVLLYGLTAATGSTSSLADYFWWIVSFSAMLLLVLSAVLARYVILLLK	60
55	orf64.pep	DRRDGVFGSXXAKXPXXXMFTLVAXLPGVFLFGFPAQFINGTINSWFGNDTHEALERSLN	120
	orf64ng	DRRNGVFGSQIAKR-LSGMFTLVAVLPGLFLFGISAQFINGTINSWFGNDTHEALERSLN	119
60	orf64.pep	LSKSALNLAADNALGNAVPVQIDLIGAASLPGDMGRVLEHYAGSGFAQLALYNXASGKIE	180
60	orf64ng	LSKSALDLAADNAVSNAVPVQIDLIGTASLSGNMGSVLEHYAGSGFAQLALYNAASGKIE	179
	orf64.pep	KSINPHKLDQPFPGKARWEKIQRAGSVRDLESIGGVLYAQGWLSAGTHXGRDYALFFRQP	240
65	orf64ng	KSINPHQFDQPLPDKEHWEQIQQTGSVRSLESIGGVLYAQGWLSAGTHNGRDYALFFRQP	239

	orf64.pep	VPKGVAEDAVLIEKARAKYAELSYSKKGLQTFFLATLLIASL		300
	orf64ng	IPENVAQDAVLIEKARAKYAELSYSKKGLQTFFLVTLLIASL		299
5	orf64.pep	EPVLSLAEGAKAVAQGDFSQTRPVLRNDEFGRLTXLFNHMTE		360
	orf64ng	EPILSLAEGAKAVAQGDFSQTRPVLRNDEFGRLTKLFNHMTE	OLSIAKEADERNRRREEA	359
10	orf64.pep	ARHYLECVLEGLTTGVVVFDEQGCLKTFNKAAGT	394	
	orf64ng	ARHYLECVLDGLTTGVVVSYPLSCCRTAVFSTCHSSPLSYF	400	

An ORF64ng nucleotide sequence <SEQ ID 255> was predicted to encode a protein having amino acid sequence <SEQ ID 256>:

```
1 MRRFLPIAAI CAVVLLYGLT AATGSTSSLA DYFWWIVSFS AMLLLVLSAV

151 LARYVILLIK DRRNGVFGSQ IAKRLSGMFT LVAVLPGLFL FGISAQFING
101 TINSWFGNDT HEALERSLNL SKSALDLAAD NAVSNAVPVQ IDLIGTASLS
151 GNMGSVLEHY AGSGFAQLAL YNAASGKIEK SINPHQFDQP LPDKEHWEQI
201 QQTGSVRSLE SIGGVLYAQG WLSAGTHNGR DYALFFRQPI PENVAQDAVL
251 IEKARAKYAE LSYSKKGLQT FFLVTLLIAS LLSIFLALVM ALYFARRFVE
20 301 PILSLAEGAK AVAQGDFSQT RPVLRNDEFG RLTKLFNHMT EQLSIAKEAD
351 ERNRREEAA RHYLECVLDG LTTGVVVSYP LSCCRTAVFS TCHSSPLSYF*
```

#### Further work revealed the complete gonococcal DNA sequence <SEQ ID 257>:

						•
	1	ATGCGCCGCT	TCCTACCGAT	CGCAGCCATA	TGCGCCGTCG	TCCTGCTGTA
	51			GCAGCACCAG		
25	101	GGTGGATAGT	CTCGTTCAGC	GCAATGCTGC	TGCTGGTGTT	GTCCGCCGTT
	151	TTGGCACGTT	ATGTCATATT	GCTGTTGAAA	GACAGGCGCA	ACGGCGTGTT
	201			GCCTTTCCGG		
	251	TACTGCCCGG	CTTGTTCCTG	TTCGGCATTT	CCGCGCAGTT	TATCAACGGC
	301	ACGATTAATT	CGTGGTTCGG	CAACGACACC	CACGAAGCCC	TCGAACGCAG
30	351	CCTTAATTTG	AGCAAGTCCG	CACTGGATTT	GGCGGCAGAC	AATGCCGTCA
	401	GCAACGCCGT	TCCCGTACAG	ATAGACCTCA	TCGGCACCGC	CTCCCTGTCG
	451	GGCAATATGG	GCAGTGTGCT	GGAACACTAC	GCCGGCAGCG	GTTTTGCCCA
	501	GCTTGCCCTG	TACAATGCCG	CAAGCGGGAA	AATCGAAAAA	AGCATCAATC
	551	CGCACCAATT	CGACCAGCCG	CTTCCCGACA	AAGAACATTG	GGAACAGATT
35	601	CAGCAGACCG	GTTCGGTTCG	GAGTTTGGAA	AGCATAGGCG	GCGTATTGTA
	651	CGCGCAGGGA	TGGTTGTCGG	CAGGTACGCA	CAACGGGCGC	GATTACGCGC
	701	TGTTCTTCCG	CCAGCCGATT	CCCGAAAATG	TGGCACAGGA	TGCCGTTCTG
	751	ATTGAAAAGG	CGCGGGCGAA	ATATGCCGAA	TTGAGTTACA	GCAAAAAAGG
	801	TTTGCAGACC	TTTTTTCTGG	TAACCCTGCT	GATTGCCTCG	CTGCTGTCGA
40	851			GCACTGTATT		
	901	CCCATTCTGT	CGCTTGCCGA	GGGCGCAAAG	GCGGTGGCGC	AGGGTGATTT
	951	CAGCCAGACG	CGCCCCGTAT	TGCGCAACGA	CGAGTTCGGA	CGTTTGACCA
	1001	AGCTGTTCAA	CCATATGACC	GAGCAGCTTT	CCATCGCCAA	AGAAGCAGAC
	1051	GAACGCAACC	GCCGGCGCGA	GGAAGCCGCC	CGTCACTACC	TCGAGTGCGT
45	1101	GTTGGATGGG	TTGACTACCG	GTGTGGTGGT	<b>GTTTGACGAA</b>	AAAGGCCGTT
	1151	TGAAAACCTT	CAACAAGGCG	GCGGAACAGA	TTTTGGGGAT	GCCGCTCGCC
	1201	CCCCTGTGGG	GCAGCAGCCG	GCACGGTTGG	CACGGCGTTT	CGGCGCAGCA
	1251	GTCCCTGCTT	GCCGAAGTGT	TtgccgccAT	CGGTGCGGCG	GCAGGTACGG
	1301	ACAAACCGGT	CCAGGTGGAA	TATGCCGCGC	CGGACGATGC	CAAAATCCTG
50	1351	CTGGGCAAGG	CGACGGTATT	GCCCGAAGAC	AACGGCAACG	GCGTGGTGAT
	1401	GGTGATTGAC	GACATCACCG	TGCTGATACG	CGCGCAAAAA	GAAGCCGCGT
	1451	GGGGTGAAGT	GGCGAAGCGG	CTGGCACACG	AAATCCGCAA	TCCGCTCACG
	.1501			ACGGCTGGCG		
	1551	GGACGATCAG	GACGCGCAAA	TCCTGACGCG	TtcgACCGAC	ACCATCATCA
55	1601	AACAGgtggc	gGCGTTAAAA	GAAATGGTCG	AGGCATTCCG	CAATTACGCG
	1651	CGCGCCCCTT	CGCTCAAACT	GGAAAATCAG	GATTTGAACG	CCTTAATCGG
	1701	CGATGTTTTG	GCCCTGTACG	AAGCCGGCCC	GTGCCGGTTT	GAGGCGGAAC
	1751	TTGCCGGCGA	ACCGCTGATG	ATGGCGGCGG	ATACGACCGC	CATGCGGCAG
	1801					AAGAAGCCGA
60	1851	TATGCCCGAA	GTCAGGGTAA	AATCGGAAAC	GGGGCAGGAC	GGACGGATTG
	1901					AATGCTGCAC
	1951					CGGGACTGGG
	2001	TCTGCCTGTA	GTGAAAAAA	TCATTGGAGA	ACACGGCGGC	CGCATCAGCC
	2051					CTTGCCAAAA
65	2101		CTTATGCGTA			

# This corresponds to the amino acid sequence <SEQ ID 258; ORF64ng-1>:

					DATEMATUCEC	VAP.TU.T.T.MA
		1 MRRFLPIAAI	CAVVLLYGLT	AATGSTSSLA	DYFWWIVSFS	WITTED A DOWA
	5	1 TADVUTTILK	DRRNGVFGSO	IAKRLSGMFT	LVAVLPGLFL	FGISAQFING
	_	1 DINEWECKIOT	HEALERSLNL	SKSALDLAAD	NAVSNAVPVQ	IDLIGTASLS
_	10		ACCCENOLAL.	YNAASGKTEK	SINPHQFDQP	LPDKEHWEQI
5	15	1 GNMGSVLEHI	MCOGEMOTIVE	THEFTOORTER	DVXI PPDADI	DEMUACDAVI.
•	20	1 OOTGSVRSLE	SIGGVLYAQG	WLSAGTHNGK	DYALFFROPI	LEMANDOMAN
	25	1 TEVADAKVAR	I.SYSKKGLOT	FFLVTLLIAS	LLSIFLALVM	ALYFARREVE
	30	1 DITTELATERK	AVAOGDESOT	RPVLRNDEFG	RLTKLFNHMT	EQLSTAKEAD
			PHYLECVING	LTTGVVVFDE	KGRLKT FNKA	AEQILGMPLA
_	35		MULTIPOLETI	AFVEDATGAA	AGTDKPVQVE	YAAPDDAKIL
10	. 40	1 PLWGSSRHGW	HPA 2WOO2TT	ADVERTION:	DARKET TRICE	TAUETDMETT
	45	1 LGKATVLPED	NGNGVVMVID	DITVLIRAQK	EAAWGEVAKR	PAUGIKAGAI
	50	1 DTOT CAFRIA	WKT.GGKLDDO	DAOILTRSTD	TIIKQVAALK	EMVEAFRNYA
	55	1 DADGIKIENO	DUNALIGDVL	ALYEAGPCRF	EAELAGEPLM	MAADTTAMRQ
			ENAFEADMPE	VRVKSETGOD	GRIVLTVCDN	GKGFGKEMLH
	60	) I APHMILKNAW	EMADEMPERO	THEFT	DISTONODAG	CACURTTI.PK
15	65	51 NAFEPYVTDK	PAGTGLGLPV	AKKTIGERGG	KISTSMODVG	GACVRIILPK
13	70	1 TVETYA*				
					- 1	
	00004 - 1 -	4 ODE64 1 d	how 93 8% i	dentity in 70	6 aa overian:	

## ORF64ng-1 and ORF64-1 show 93.8% identity in 706 aa overlap:

	OKPO411g-1 and Older	J4-1 BIIOW > DIG / U			•		
	orf64ng-1.pep	10 MRRFLPIAAICAVVL	20 LYGLTAATG	30 STSSLADYFW	40 NIVSFSAMLLI	50 LVLSAVLARYV	60 VILLLK
20	Officing 1.pop			111111111	! [   :	1	
20	orf64-1	MRRFLPIAAICAVVL	LYGLTAATG	STSSLADYEW	MIVALSAMPPI	PATOWATIVE A	60 TTTTK
	02200	10	20	30	40	50	60
				00	100	110	120
		70 DRRNGVFGSQIAKRL	80	90			
25	orf64ng-1.pep	DRRNGVFGSQIAKRL	SGMFTLVAV	1111111111	1111111111		111111
		DRRDGVFGSQIAKRL	I I I I I I I I I I I I I I I I I I I	LPGVFLFGVS	AOFINGTINS	WFGNDTHEALI	ERSLNL
	orf64-1	70	80	90	100	110	120
		,,	**				
30		130	140	150	160	170	180
30	orf64ng-1.pep	SKSALDLAADNAVSN	AVPVQIDL	[GTASLSGNMG	SVLEHYAGSG	FAQLALYNAA	SGKIEK
	02203-1-1			11:11 1:11		111111111	,,,,,,
	orf64-1	SKSALNLAADNALGN	AVBAÖLDT:	IGAASLPGDMG	RVLEHYAGSG 160	170	180
		130	140	150	100	170	
35		190	200	210	220	230	240
		CTARROCK OF POKE	тоотоями	GSVRSLESIGO	VLYAOGWLSA	GTHNGRDYAL	FFRQPI
	orf64ng-1.pep			1111211111			11111
	orf64-1	SINPHKLDQPFPGK/	ARWEKIORA	GSVRDLESIGO	VLYAQGWLSA	GTHNGKDIAL	E L L L L L L L
40	01104-1	190	200	210	220	230	240
40						222	300
		250	260	270	280	290	
	orf64ng-1.pep	PENVAQDAVLIEKA	RAKYAELSY	SKKGLQTFFL	/TLLIASLLS1		IIIIIII
		::  :           PKGVAEDAVLIEKA				:	ARREVE
45	orf64-1		RAKYAELSI 260	270	280	290	300
		250	200	210	200		
		310	320	330	340	350	360
	orf64ng-1.pep	DIT OF APCAKANAO	CDESCUEPY	LRNDEFGRLT	KLFNHMTEQL:	SIAKEADERNI	RRREEAA
50	Offgang-1.beb		111111111	111111111		1111111	, , , , , , ,
30	orf64-1	PVLSLAEGAKAVAQ	GDFSQTRPV	LRNDEFGRLT	KTENHWLEGT:	STAKEADEKM	RRREEAA 360
	32233	310	320	330	340	350	360
			200	390	400	410	420
		370 RHYLECVLDGLTTG	380	> 7 50 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	TT.CMPT.APT.W		
55	orf64ng-1.pep	RHYLECVLDGLTTG	5000EDEKG	111111111	111111:11	1111111111	1111111
	ec. 1	RHYLECVLEGLTTG	VVVFDEOG	LKTFNKAAEO	ILGMPLTPLW	GSSRHGWHGV	SAQQSLL
	orf64-1	370	380	390	400	410	420
		3.0					
60		430	440	450	460	470	480
00	orf64ng-1.pep	AEVFAAIGAAAGTI	DKPVQVEYA	APDDAKILLGK	ATVLPEDNGN	GAAWAIDDII	VLIRAQK
	Quality of Pop			111111111	11111111111		111:111
	orf64-1	AEVFAAIGAAAGT		APDDAKILLG	ATV LPEDNGN	470	480
		430	440	450	460	4,0	400
65		490	500	510	520	530	540
		TARKCEUNKOT AH	ETRNPLTPT	OT.SAERT.AWK	LGGKLDDQDAQ	OILTRSTDTII	KQVAALK
	orf64ng-1.peg	EAAWGEVARRAH 	711111111 77045 DIET	11111111111			:::::::::::::::::::::::::::::::::::::::
		[[]]]]					

```
orf64-1
                        EAAWGEVAKRLAHEIRNPLTPIQLSAERLAWKLGGKLDEQDAQILTRSTDTIVKOVAALK
                              490
                                        500
                                                 510
                                                           520
                              550
                                        560
                                                 570
                                                           580
5
          orf64ng-1.pep
                        EMVEAFRNYARAPSLKLENQDLNALIGDVLALYEAGPCRFEAELAGEPLMMAADTTAMRO
                        EMVEAFRNYARSPSLKLENQDLNALIGDVLALYEAGPCRFAAELAGEPLTVAADTTAMRQ
          orf64-1
                                        560
                                                 570
                                                           580 ·
10
                              610
                                        620
                                                 630
                                                           640
                                                                    650
                                                                             660
                        VLHNIFKNAAEAAEEADMPEVRVKSETGQDGRIVLTVCDNGKGFGKEMLHNAFEPYVTDK
          orf64ng-1.pep
                        orf64-1
                        VLHNIFKNAAEAAEEADVPEVRVKSETGQDGRIVLTVCDNGKGFGREMLHNAFEPYVTDK
                              610
                                        620
                                                 630
                                                           640
                                                                    650
15
                              670
                                        680
                                                 690
                                                          . 700
                        PAGTGLGLPVVKKIIGEHGGRISLSNQDAGGACVRIILPKTVETYAX
          orf64ng-1.pep
                        orf64-1
                        PAGTGLGLPVVKKIIEEHGGRISLSNQDAGGACVRIILPKTVKTYAX
20
                                        680
                                                 690
                                                           700
     Furthermore, ORF64ng-1 shows significant homology to a protein from A. caulinodans:
          sp|Q04850|NTRY AZOCA NITROGEN REGULATION PROTEIN NTRY >gi|77479|pir||S18624 ntry
          protein - Azorhizobium caulinodans >gi|38737 (X63841) NtrY gene product
          [Azorhizobium caulinodans] Length = 771
25
           Score = 218 bits (550), Expect = 7e-56
           Identities = 195/720 (27%), Positives = 320/720 (44%), Gaps = 58/720 (8%)
                    IAAICAVVLLYGLTAATGSTSSLADYFWWIXXXXXXXXXXXXXXXXXXXXXXXXXXX 66
          Query: 7
                    I+A+
                          ++L GLT
30
          Sbjct: 35 ISALATFLILMGLTPVVPTHQVVIS----VLLVNAAAVLILSAMVGREIWRIAKARARGR 90
          Query: 67
                    FGSQIAKRLSGMFTLVAVLPGLFLFGISAQFINGTINSWFGNDTHEALERSLNLSKSALD 126
                      +++ R+ G+F +V+V+P + + +++ ++ WF
                                                           TE + S++++++ +
          Sbjct: 91 AAARLHIRIVGLFAVVSVVPAILVAVVASLTLDRGLDRWFSMRTQEIVASSVSVAQTYVR 150
35
          Query: 127 LAADNAVSNAVPVQIDLIGTASLSGNMGSVLEHYAG--SGFAQLALYNAASGKIEKSINP 184
                      A N + + + DL
                                        S±
                                                   YGSFQ+
                                                                 AA
          Sbjct: 151 EHALNIRGDILAMSADLTRLKSV-
                                                 --YEGDRSRFNQILTAQAALRNLPGAMLI 200
40
          Query: 185 HQFDQPLPDKEHWEQIQQTGSVRSLESIGGVLYAQGWLSAGTHNGRDYA----- 233
                     + D + ++ + I + V + +IG
                                                  Q +
          Sbjct: 201 RR-DLSVVERAN-VNIGREFIVPANLAIGDATPDQPVIYLP--NDADYVAAVVPLKDYDD 256
          Query: 234 --LFFRQPIPENVAQDAVLIEKARAKYAELSYSKKGLQTFFLVTXXXXXXXXXXXXXXXXXX 291
45
                      L+ + I V
                                      ++ A Y L
                                                  + G+Q F +
          Sbjct: 257 LYLYVARLIDPRVIGYLKTTQETLADYRSLEERRFGVQVAFALMYAVITLIVLLSAVWLG 316
          Query: 292 LYFARRFVEPILSLAEGAKAVAQGDFSQTRPVLRND-EFGRLTKLFNHMTEQLSIXXXXX 350
                    L F++ V PI L A VA+G+
                                               P+ R + + L + FN MT +L
50
          Sbjct: 317 LNFSKWLVAPIRRLMSAADHVAEGNLDVRVPIYRAEGDLASLAETFNKMTHELRSQREAI 376
          Query: 351 XXXXXXXXXXXYLECVLDGLTTGVVVFDEKGRLKTFNKAAEQILGMPLAPLWGSSRHGW 410
                               + E VL G+ GV+ D + R+
                                                     N++AE++LG L+ +
          Sbjct: 377 LTARDQIDSRRRFTEAVLSGVGAGVIGLDSQERITILNRSAERLLG--LSEVEALHRHLA 434
55
          Query: 411 HGVSAQQSLLAEVFXXXXXXXXTDKPVQVEYAAPDDAKILLGKATVLPEDNG---NGVVM 467
                           LL E
                                          + VQ
                                                     D + +
                                                             VE+
                                                                        +G V+
          Sbjct: 435 EVVPETAGLLEEA-----EHARQRSVQGNITLTRDGRERVFAVRVTTEQSPEAEHGWVV 488
60
          Query: 468 VIDDITVLIRAQKEAAWGEVAKRLAHEIRNPLTPIQLSAERLAWKLGGKLDDQDAQILTR 527
                     +DDIT LI AQ+ +AW +VA+R+AHEI+NPLTPIQLSAERL K G + QD +I +
          Sbjct: 489 TLDDITELISAQRTSAWADVARRIAHEIKNPLTPIQLSAERLKRKFGRHV-TQDREIFDQ 547
          Query: 528 STDTIIKQVAALKEMVEAFRNYARAPSLKLENQDLNALIGDVLALYEAGPCRFEAELAGE 587
65
                     TDTII+QV + MV+ F ++AR P
                                              +++QD++ +I
                                                          + L
          Sbjct: 548 CTDTIIRQVGDIGRMVDEFSSFARMPKPVVDSQDMSEIIRQTVFLMRVGHPEVVFDSEVP 607
          Query: 588 PLMMAA-DTTAMRQVLHNIFKNXXXXXXXXDMPEVRVK-
                                                              -SETGQDGRIVLTVCD 639
                    PMAD + QLNIKN
                                                  P+VR +
                                                              + G+D +V+ + D
70
          Sbjct: 608 PAMPARFDRRLVSQALTNILKNAAEAIEAVP-PDVRGQGRIRVSANRVGED--LVIDIID 664
```

```
Query: 640 NGKGFGKEMLHNAFEPYVTDKPAGTGLGLPVVKKIIGEHGGRISLSNQDAG-GACVRIIL 698
NG G +E + EPYVT + GTGLGL +V KI+ EHGG I L++ G GA +R+ L
Sbjct: 665 NGTGLPQESRNRLLEPYVTTREKGTGLGLAIVGKIMEEHGGGIELNDAPEGRGAWIRLTL 724
```

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### 10 Example 31

5

The following partial DNA sequence was identified in N.meningitidis <SEQ ID 259>:

```
1 ATGTACGCAT TTACCGCCGC ACAGCAACAG AAGGCACTCT TCCGGCTGGT
                    GCTTTTTCAT ATCCTCATCA TCGCCGCCAG CAACTATCTG GTGCAGTTCC
                    CTTTCCAAAT TTTCGGCATC CACACCACTT GGGGCGCATT TTCCTTTCCC
                101
                    TTCATCTTCC TTGCCACCGA CCTGACCGTC CGCATTTTCG GTTCTCACTT
15
                    GGCACGGCGG ATTATCTTTT GGGTGATGTT CCCCGCCCTT TTGCTTTCCT
                201
                251 ACGTCTTTTC CGTTTTGTTC CACAACGGCA GTTGGACAGG CTTGGGCGCG
                    CTGTCCGAAT TCAACACCTT TGTCGGACGC ATCGCCTTAG CCAGCTTTGC
                301
                351 CGCCTACGCG ATCGGACAAA TCCTTGATAT TTTTGTATTC AACAAATTAC
                     GCCGTCTGAA AGCGTGGTGG ATTGCACCGA ACGCATCAAC CGTCATCGGG
                401
20
                     CACGCGTTGG ATACG...
                451
```

This corresponds to the amino acid sequence <SEQ ID 260; ORF66>:

```
25 MYAFTAAQQQ KALFRLVLFH ILIIAASNYL VQFPFQIFGI HTTWGAFSFP
FIFLATDLTV RIFGSHLARR IIFWVMFPAL LLSYVFSVLF HNGSWTGLGA
LSEFNTFVGR IALASFAAYA IGQILDIFVF NKLRRLKAWW IAPNASTVIG
```

Further work revealed the complete nucleotide sequence <SEQ ID 261>:

```
1 ATGTACGCAT TTACCGCCGC ACAGCAACAG AAGGCACTCT TCCGGCTGGT
                    GCTTTTCAT ATCCTCATCA TCGCCGCCAG CAACTATCTG GTGCAGTTCC
                51
                    CTTTCCAAAT TTTCGGCATC CACACCACTT GGGGCGCATT TTCCTTTCCC
30
               101
               151 TTCATCTTCC TTGCCACCGA CCTGACCGTC CGCATTTTCG GTTCTCACTT
                    GGCACGGCGG ATTATCTTTT GGGTGATGTT CCCCGCCCTT TTGCTTTCCT
               201
                    ACGTCTTTTC CGTTTTGTTC CACAACGGCA GTTGGACAGG CTTGGGCGCG
                251
                    CTGTCCGAAT TCAACACCTT TGTCGGACGC ATCGCCTTAG CCAGCTTTGC
                301
                    CGCCTACGCG ATCGGACAAA TCCTTGATAT TTTTGTATTC AACAAATTAC
35
                351
                    GCCGTCTGAA AGCGTGGTGG ATTGCACCGA CCGCATCAAC CGTCATCGGC
                401
                    AACGCCTTGG ATACGCTGGT ATTTTTCGCC GTTGCCTTCT ACGCAAGCAG
                451
                    CGATGGATTT ATGGCGGCAA ACTGGCAGGG CATCGCTTTT GTCGATTACC
                501
                    TGTTCAAACT TACCGTCTGC ACCCTCTTCT TCCTGCCCGC CTACGGCGTG
                551
                    ATACTGAATC TGCTGACGAA AAAACTGACA ACCCTGCAAA CCAAACAGGC
                601
40
                    GCAAGACCGC CCCGCGCCCT CGCTGCAAAA TCCGTAA
```

This corresponds to the amino acid sequence <SEQ ID 262; ORF66-1>:

```
45 MYAFTAAQQQ KALFRLVLFH ILIIAASNYL VQFPFQIFGI HTTWGAFSFP
FIFLATDLTV RIFGSHLARR IIFWVMFPAL LLSYVFSVLF HNGSWTGLGA
LSEFNTFVGR IALASFAAYA IGQILDIFVF NKLRRLKAWW IAPTASTVIG
151 NALDTLVFFA VAFYASSDGF MAANWQGIAF VDYLFKLTVC TLFFLPAYGV
201 ILNLLTKKLT TLQTKQAQDR PAPSLQNP*
```

Computer analysis of this amino acid sequence gave the following results:

Homology with the hypothetical protein o221 of E. coli (accession number P37619)

ORF66 and o221 protein show 67% aa identity in 155aa overlap:

		-109-
	orf66 1	MYAFTAAQQQKALFRLVLFHILIIAASNYLVQFPFQIFGIHTTWGAFSFPFIFLATDLTV 60 M F+ Q+ KALF L LFH+L+I +SNYLVQ P I G HTTWGAFSFPFIFLATDLTV
	0221 1	MNVFSQTQRYKALFWLSLFHLLVITSSNYLVQLPVSILGFHTTWGAFSFPFIFLATDLTV 60
5	orf66 61	RIFG+ LARRIIF VM PALL+SYV S LF+ GSW G GAL+ FN FV RIA ASF AYA
	0221 61	RIFGAPLARRIIFAVMIPALLISYVISSLFYMGSWQGFGALAHFNLFVARIATASFMAYA 120
10		1 IGQILDIFVFNKLRRLKAWWIAPNASTVIGHALDT 155 +GQILD+ VFN+LR+ + WW+AP AST+ G+ DT
		1 LGQILDVHVFNRLRQSRRWWLAPTASTLFGNVSDT 155
		a predicted ORF from N.meningitidis (strain A)
	ORF66 shows 9	6.1% identity over a 155aa overlap with an ORF (ORF66a) from strain A of N.
15	meningitidis:	
	orf66.pep	10 20 30 40 50 60 MYAFTAAQQQKALFRLVLFHILIIAASNYLVQFPFQIFGIHTTWGAFSFPFIFLATDLTV
20	orf66a	MYAFTAAQQQKALFWLVLFHILIIAASNYLVQFPFQISGIHTTWGAFSFPFIFLATDLTV 10 20 30 40 50 60
	orf66.pep	70 80 90 100 110 120 RIFGSHLARRIIFWVMFPALLLSYVFSVLFHNGSWTGLGALSEFNTFVGRIALASFAAYA
25	orf66a	
23	Offeda	RIFGSHLARRIIFWVMFPALLLSYVFSVLFHNGSWTGLGALSEFNTFVGRIALASFAAYA 70 80 90 100 110 120
	orf66.pep	130 140 150 IGQILDIFVFNKLRRLKAWWIAPNASTVIGHALDT
30	orf66a	:
		130 140 150 160 170 180
35	orf66a	VDYLFKLT <u>VCGLFFLPAYGVILNLL</u> TKKLTTLQTKQAQDRPAPSLQNPX 190
	The complete le	ngth ORF66a nucleotide sequence <seq 263="" id=""> is:</seq>
	1	ATGTACGCAT TTACCGCCGC ACAGCAACAG AAGGCACTCT TCTGGCTGGT
	51 101	GCTTTTTCAT ATCCTCATCA TCGCCGCCAG CAACTATCTG GTGCAGTTCC CCTTCCAAAT TTCCGGCATC CACACCACTT GGGGCGCGTT TTCCTTTCCC
40	151	TTCATCTTCC TCGCCACCGA CCTGACCGTC CGCATTTTCG GTTCGCACTT
	201 251	GGCACGGCGG ATTATCTTTT GGGTCATGTT CCCCGCCCTT TTGCTTTCCT ACGTCTTTTC CGTTTTGTTC CACAACGGCA GTTGGACGGG CTTGGGCCGCG
	301 351	CTGTCCGAAT TCAACACCTT TGTCGGACGC ATCGCGCTGG CAAGTTTTGC CGCCTACGCG CTCGGACAAA TCCTTGATAT TTTTGTGTTC AACAAATTAC
45	401	GCCGTCTGAA AGCGTGGTGG GTTGCCCCGA CTGCATCAAC CGTCATCGGC
	451 501	AACGCCTTAG ATACGTTGGT ATTTTTCGCC GTTGCCTTCT ACGCAAGCAG CGATGGATTT ATGGCGGCAA ACTGGCAGGG CATCGCTTTT GTCGATTACC
	551	TGTTCAAACT CACCGTCTGC GGTCTGTTTT TCCTGCCCGC CTACGGCGTG
50	601 651	ATTCTGAATC TGCTGACGAA AAAACTGACG ACCCTGCAAA CCAAACAGGC GCAAGACCGC CCCGCGCCCT CGCTGCAAAA TCCGTAA
	This encodes a p	protein having amino acid sequence <seq 264="" id="">:</seq>
	1	MYAFTAAQQQ KALFWLVLFH ILIIAASNYL VQFPFQISGI HTTWGAFSFP
	51	FIFLATDLTV RIFGSHLARR <u>IIFWVMFPAL</u> LLSYVFSVLF HNGSWTGLGA
55	101 151	LSEFNTFVGR IALASFAAYA LGQILDIFVF NKLRRLKAWW VAPTASTVIG NALDTLVFFA VAFYASSDGF MAANWQGIAF VDYLFKLTVC GLFFLPAYGV
	201	ILNLLTKKLT TLQTKQAQDR PAPSLQNP*
	ORF66a and OF	RF66-1 show 97.8% identity in 228 aa overlap:
	orf66a.pe	10 20 30 40 50 60 p MYAFTAAQQQKALFWLVLFHILIIAASNYLVQFPFQISGIHTTWGAFSFPFIFLATDLTV
60	_	
	orf66-1	MYAFTAAQQQKALFRLVLFHILIIAASNYLVQFPFQIFGIHTTWGAFSFPFIFLATDLTV

						PCT/I	B98/01665
	WO 99/24578		-1	90-			
				-	40	50	60
		10	20	30	40	50	
		70	80	90	100	110	120 FAAYA
	orf66a.pep	70 RIFGSHLARRIIFWVM					
5	orf66-1		IFPALLLSY	AF2AFF UNGS	MIGDOMEDED.	ENTEVGRIALASI 110	FAAYA 120
	01100-1	70	80	90	100	110	
	,	130	140	150	160	170	180 OGTAF
10	orf66a.pep	130 LGQILDIFVFNKLRRI :					
	orf66-1	:	PKAMMIWE.	LW21AIGNWDI	77 TA T T11411	YASSDGFMAANW 170	QGIAF 180
	01200 -	130	140	150	160	2.0	
15		190	200	210	220	229	
13	orf66a.pep	VDYLFKLTVCGLFFL	1111111	1 1 1 2 1 1 1 1 1 1	111111111	11111	
	orf66-1	VDYLFKLTVCTLFFL	PAYGVILN	PPIKKPILITO	TKQAQDRPAF 220	SLQNPX	
	<b>42</b>	190	200	210	220		
20	Homology with a	predicted ORF from	n <i>N.gono</i>	<u>rrhoeae</u>			
	HOMOTORY WILL	2% identity over a	15599.03	erlan with	a predicted	ORF (ORF	66.ng) from N.
	ORF66shows 94.	2% identity over a	133aa Ov	Oliup Willi	w pro	•	
	gonorrhoeae:						
		MYAFTAAQQQKALFI	RLVLFHIL!	IJASNYLVQE	PFQIFGIHT	TWGAFSFPFIFL	ATDLTV 60
25	orf66.pep	MYAFTAAQQQKALFI    :         MYALTAAQQQKALFI					
20	orf66ng						
	orf66.pep	RIFGSHLARRIIFW					
30	orf66ng		VMFPALSL	SYVFSVLFHN	GSWTGLGAPS	QFNTFVGRIALA	SFAAYA 120
30	-	IGQILDIFVFNKLR					155
	orf66.pep					LEVA SSDEFMAAN	WOGIAF 180
	orf66ng	<b>LGQILDIFVFDKL</b> R	RLKAWWIA	PAASTVIGNA	TOIPALLWAL		
35	The complete ler	ngth ORF66ng nucl	eotide se	quence <se< td=""><td>Q ID 265&gt;</td><td>· 1S:</td><td></td></se<>	Q ID 265>	· 1S:	
					GGCACTCT T	rCCGGCTGGT	
			33 max max	TOTAL CA	MILLAILIG	3100001100	
		CCTTCCGGAT TTTCGG	ACCCA CCC	PGACCGTC CO	CATTTTCG	GTTCGCACTT	
40	151 201		- <b>ጥጥጥጥ ርር</b> ር	THEATERIN CO		ccgc+++	
40	251		<b>"""""""""""""""""""""""""""""""""""""</b>	"AA("-(-) A (-)	UUUUUUU	011000000	
	301		ויוי מאאיאא	CTITICATIATE I.	IICGIDIIC	O* TO:	
	351 401						
45	451		ጥጥ እርጥ አጥ	TTTTTTTCCCC G	TIGULLIA	MOOGRAGOISE	
75	501	CGATGAATTT ATGGC					
	551	አመአረመሮአአጥሮ ጥርሮጥር	ACGAA AA	AACTGACG G	CCCIGCIA	CCAAACAGGC	
	601 651	GCAAGACCGC CCCGT	GCCCT CG	CTGCAAAA T	CCGTAA		
<b>5</b> 0	This encodes a	protein having amir	no acid se	equence <si< td=""><td>EQ ID 266</td><td>&gt;:</td><td></td></si<>	EQ ID 266	>:	
50							
		MYALTAAQQQ KALFF FIFLATDLTV RIFGS	וו ססגדנוי	I WWW PAL 3	TOTALDAME	111100112-0-1	
	51 101						
	151	NATIOTIVEFA VAFYA	ASSDEF M	AANWQGIAF V	DYLFKLTVC	TLFFLPAYGV	
55	201	ILNLLTKKLT ALOTE	KOAQDR P	^ 527∂N5 <sub>*</sub>			

An alternative annotated sequence is:

ORF66ng and ORF66-1 show 96.1% identity in 228 aa overlap:

```
orf66-1.pep
                MYAFTAAQQQKALFRLVLFHILIIAASNYLVQFPFQIFGIHTTWGAFSFPFIFLATDLTV
                 orf66ng
                 MYALTAAQQQKALFRLVLFHILIIAASNYLVQFPFRIFGIHTTWGAFSFPFIFLATDLTV
5
       orf66-1.pep
                 RIFGSHLARRIIFWVMFPALLLSYVFSVLFHNGSWTGLGALSEFNTFVGRIALASFAAYA 120
                 RIFGSHLARRIIFWVMFPALLLSYVFSVLFHNGSWTGLGALSQFNTFVGRIALASFAAYA 120
       orf66ng
10
                 IGQILDIFVFNKLRRLKAWWIAPTASTVIGNALDTLVFFAVAFYASSDGFMAANWQGIAF 180
       orf66-1.pep
                 LGQILDIFVFDKLRRLKAWWIAPAASTVIGNALDTLVFFAVAFYASSDEFMAANWQGIAF 180
       orf66ng
       orf66-1.pep
                 VDYLFKLTVCTLFFLPAYGVILNLLTKKLTTLQTKQAQDRPAPSLQNPX
                                                      229
15
                 orf66ng
                 VDYLFKLTVCTLFFLPAYGVILNLLTKKLTALQTKQAQDRPVPSLQNPX
                                                      229
```

Furthermore, ORF66ng shows significant homology with an E.coli ORF:

```
sp|P37619|YHHQ ECOLI HYPOTHETICAL 25.3 KD PROTEIN IN FTSY-NIKA INTERGENIC
          REGION (0221)
20
          >gi|1073495|pir||S47690 hypothetical protein o221 - Escherichia coli >gi|466607
           (U00039) No definition line found [Escherichia coli] >gi|1789882 (AE000423)
          hypothetical 25.3 kD protein in ftsY-nikA intergenic region [Escherichia coli]
          Length = 221
            Score = 273 bits (692), Expect = 5e-73
25
            Identities = 132/203 (65%), Positives = 155/203 (76%)
                     {\tt MYALTAAQQQKALFRLVLFHILIIAASNYLVQFPFRIFGIHTTWGAFSFPFIFLATDLTV~60}
          Query: 1
                          + Q+ KALF L LFH+L+I +SNYLVQ P I G HTTWGAFSFPFIFLATDLTV
          Sbjct: 1
                      MNVFSQTQRYKALFWLSLFHLLVITSSNYLVQLPVSILGFHTTWGAFSFPFIFLATDLTV 60
30
                     RIFGSHLARRIIFWVMFPALLLSYVFSVLFHNGSWTGLGALSQFNTFVGRIALASFAAYA 120
          Query: 61
                      RIFG+ LARRIIF VM PALL+SYV S LF+ GSW G GAL+ FN FV RIA ASF AYA
                     RIFGAPLARRIIFAVMIPALLISYVISSLFYMGSWQGFGALAHFNLFVARIATASFMAYA 120
           Sbict: 61
35
          Query: 121 LGQILDIFVFDKLRRLKAWWIAPAASTVIGNALDTLVFFAVAFYASSDEFMAANWQGIAF 180
                      LGQILD+ VF++LR+ + WW+AP AST+ GN DTL FF +AF+ S D FMA +W IA
           Sbjct: 121 LGQILDVHVFNRLRQSRRWWLAPTASTLFGNVSDTLAFFFIAFWRSPDAFMAEHWMEIAL 180
           Query: 181 VDYLFKLTVCTLFFLPAYGVILN 203
40
                      VDY FK+ + +FFLP YGV+LN
           Sbjct: 181 VDYCFKVLISIVFFLPMYGVLLN 203
```

Based on this analysis, including the homology with the *E.coli* protein and the presence of several putative transmembrane domains in the gonococcal protein, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 32

45

The following partial DNA sequence was identified in N.meningitidis <SEO ID 267>:

	1	ATGGTCATAA	AATATACAAA	TTTGAATTTT	GCGAAATTGT	CGATAATTGC
50	51	AATTTTGATG	ATGTATTCGT	TTGAAGCGAA	TGCAAAyGCA	GTmwrAATAT
	101				GTGCGAAAAT	
	151				GATTTAATAA	
	201				CCGAATCAAC	
	251					CAAACTTGCC
55	301					TCGGAACAGC
	351				CAAAGAAGAC	
	401	GAGGCTACCA	ATACGACCCC	GAAACCGACA	AATTTGTAAA	AGGCTACGAA
	451	TATAGTAATT	GCCTTTGGTA	CGAAGACAAA	<b>AGACGTATTA</b>	ATAGAACCTA

501 TGGCTGCTAC GGCGTTGAT..

This corresponds to the amino acid sequence <SEQ ID 268; ORF72>:

```
1 MVIKYTNLNF AKLSIIAILM MYSFEANANA VXISETVSVD TGQGAKIHKF
51 VPKNSKTYSS DLIKTVDLTH XPTGAKARIN AKITASVSRA GVLAGVGKLA
5 101 RLGAKFSTRA VPYVGTALLA HDVYETFKED IQARGYQYDP ETDKFVKGYE
151 YSNCLWYEDK RRINRTYGCY GVD..
```

Further work revealed the complete nucleotide sequence <SEQ ID 269>:

```
1 ATGGTCATAA AATATACAAA TTTGAATTTT GCGAAATTGT CGATAATTGC
                    AATTTTGATG ATGTATTCGT TTGAAGCGAA TGCAAATGCA GTAAAAATAT
                    CTGAAACTGT TTCAGTTGAT ACCGGACAAG GTGCGAAAAT TCATAAGTTT
10
               101
               151 GTACCTAAAA ATAGTAAAAC TTATTCATCT GATTTAATAA AAACGGTAGA
               201 TTTAACACAC ATCCCTACGG GCGCAAAAGC CCGAATCAAC GCCAAAATAA
                    CCGCCAGCGT ATCCCGCGCC GGCGTATTGG CGGGGGTCGG CAAACTTGCC
               251
                    CGCTTAGGCG CGAAATTCAG CACAAGGGCG GTTCCCTATG TCGGAACAGC
               301
                    CCTTTTAGCC CACGACGTAT ACGAAACTTT CAAAGAAGAC ATACAGGCAC
               351
15
                    GAGGCTACCA ATACGACCCC GAAACCGACA AATTTGCAAA GGTCTCAGGC
                401
                451
```

This corresponds to the amino acid sequence <SEQ ID 270; ORF72-1>:

```
20 51 MVIKYTNLNF AKLSIIAILM MYSFEANANA VKISETVSVD TGQGAKIHKF
VPKNSKTYSS DLIKTVDLTH IPTGAKARIN AKITASVSRA GVLAGVGKLA
101 RLGAKFSTRA VPYVGTALLA HDVYETFKED IQARGYQYDP ETDKFAKVSG
151 *
```

Computer analysis of this amino acid sequence gave the following results:

## Homology with a predicted ORF from N.meningitidis (strain A)

ORF72 shows 98.0% identity over a 147aa overlap with an ORF (ORF72a) from strain A of N. meningitidis:

		10	20	30	40	50	60
	orf72.pep	MVIKYTNLNFAKL	SIIAILMMYSF	<u>EANA</u> NAVXIS	SETVSVDTGQG	AKIHKEVPKN	ISKTYSS
	022.012.2	111111111111		311111111111		1111111111	111111
30	orf72a	MVIKYTNLNFAKL	SIIAILMMYSF	<u>EANA</u> NAVKIS	SETVSVDTGQG	AKIHKEVPKN	ISKTYSS
30	<b>0.2</b> ·	10	20	30	40	50	60
		70	80	90	100	110	120
	<b>450</b>	DLIKTVDLTHXPT	OU CAVADINAVIT				GTALLA
0.5	orf72.pep	DUTKLADDITUKET	GWWWINWII	111111111		1111111111	ШШ
35	orf72a	DLIKTVDLTHIPT	GAKARINAKIT	ASVSRAGVL	AGVGKLARLGA	KFSTRAVPYV	/GTALLA
	OII/Za	70	80	90	100	110	120
				150	160	170	
		130	140	150			
40	orf72.pep	HDVYETFKEDIQA			CTMIFDVKKID	WIIGCIGAD	
		_		:			
	orf72a	HDVYETFKEDIQA		KFAKVSGX			
		130	140	150			

The complete length ORF72a nucleotide sequence <SEQ ID 271> is:

	<del>-</del>					
45	1	ATGGTCATAA	AATATACAAA	TTTGAATTTT	GCGAAATTGT	CGATAATTGC
73	51	AATTTTGATG	ATGTATTCGT	TTGAAGCGAA	TGCAAATGCA	GTAAAAATAT
	101	CTGAAACTGT	TTCAGTTGAT	ACCGGACAAG	GTGCGAAAAT	TCATAAGTTT
	151	GTACCTAAAA	ATAGTAAAAC	TTATTCATCT	GATTTAATAA	AAACGGTAGA
	201	TTTAACACAC	ATCCCTACGG	GCGCAAAAGC	CCGAATCAAC	GCCAAAATAA
50	251	CCGCCAGCGT	ATCCCGCGCC	GGCGTATTGG	CGGGGGTCGG	CAAACTTGCC
30	301	CGCTTAGGCG	CGAAATTCAG	CACAAGGGCG	GTTCCCTATG	TCGGAACAGC
	351	CCTTTTAGCC	CACGACGTAT	ACGAAACTTT	CAAAGAAGAC	ATACAGGCAC
	401	GAGGCTACCA	ATACGACCCC	GAAACCGACA	AATTTGCAAA	GGTCTCAGGC
	451	TAA	*·			
	401					

55 This encodes a protein having amino acid sequence <SEQ ID 272>:

1	MVIKYTNLNF	AKLSIIAILM	MYSFEANANA	VKISETVSVD	TGQGAKIHKF
51	VPKNSKTYSS	DLIKTVDLTH	IPTGAKARIN	AKITASVSRA	GVLAGVGKLA
101	RLGAKFSTRA	VPYVGTALLA	HDVYETFKED	IQARGYQYDP	ETDKFAKVSG
151	*	••			

5 ORF72a and ORF72-1 show 100.0% identity in 150 aa overlap:

		10	20	30	40	50	60
	orf72a.pep	MVIKYTNLNFAKL:	SIIAILMMYSE	EANANAVKIS	SETVSVDTGQ0	Sakihkevpki	NSKTYSS
	orf72-1	MVIKYTNLNFAKL	SIIAILMMYSF	EANANAVKIS	SETVSVDTGQ(	Sakihkevpki	NSKTYSS
10		10	20	30	40	50	60
		70	80	, 90	100	110	120
	orf72a.pep	DLIKTVDLTHIPT	GAKARINAKIT	ASVSRAGVL	AGVGKLARLG!	KFSTRAVPY	
			111111111111	11111111	11111111		111111
15	orf72-1	DLIKTVDLTHIPT					
		70	80	90	100	110	120
		130	140	150	•		
	orf72a.pep	HDVYETFKEDIQA	RGYQYDPETDK	FAKVSGX			
20			111111111111	111111			•
	orf72-1	HDVYETFKEDIQA	RGYQYDPETDI	FAKVSGX			
		130	140	150			

### Homology with a predicted ORF from N.gonorrhoeae

ORF72 shows 89% identity over a 173aa overlap with a predicted ORF (ORF72.ng) from N.

```
gonorrhoeae:
                                                               60
       orf72.pep
                 MVIKYTNLNFAKLSIIAILMMYSFEANANAVXISETVSVDTGQGAKIHKFVPKNSKTYSS
                 MVTKHTNLNFAKLSIIAILMMYSFEANANAVKISETLSVDTGQGAKVHKFVPKSSNIYSS
                                                               60
       orf72ng
30
       orf72.pep
                 DLIKTVDLTHXPTGAKARINAKITASVSRAGVLAGVGKLARLGAKFSTRAVPYVGTALLA
                                                               120
                 DLTKAVDLTHIPTGAKARINAKITASVSRAGVLSGVGKLVRQGAKFGTRAVPYVGTALLA
                                                              120
       orf72ng
35
       orf72.pep
                 HDVYETFKEDIQARGYQYDPETDKFVKGYEYSNCLWYEDKRRINRTYGCYGVD
                                                               173
                 HDVYETFKEDIQARGCRYDPETDKFVKGYEYANCLWYEDERRINRTYGCYGVDSSIMRLM
       orf72ng
```

An ORF72ng nucleotide sequence <SEQ ID 273> was predicted to encode a protein having amino acid sequence <SEQ ID 274>:

```
40
                    MVTKHTNLNF AKLSIIAILM MYSFEANANA VKISETLSVD TGQGAKVHKF
                51
                    VPKSSNIYSS DLTKAVDLTH IPTGAKARIN AKITASVSRA GVLSGVGKLV
                    RQGAKFGTRA VPYVGTALLA HDVYETFKED IQARGCRYDP ETDKFVKGYE
               101
                    YANCLWYEDE RRINRTYGCY GVDSSIMRLM PDRSRFPEVK QLMESQMYRL
               151
               201 ARPFWNWRKE ELNKLSSLDW NNFVLNRCTF DWNGGGCAVN KGDDFRAGAS
45
               251 FSLGRNPKYK EEMDAKKPEE ILSLKVDADP DKYIEATGYP GYSEKVEVAP
               301 GTKVNMGPVT DRNGNPVQVA ATFGRDAQGN TTADVQVIPR PDLTPASAEA
                    PHAOPLPEVS PAENPANNPD PDENPGTRPN PEPDPDLNPD ANPDTDGQPG
               351
                    TSPDSPAVPD RPNGRHRKER KEGEDGGLSC DYFPEILACQ EMGKPSDRMF
               401
               451
                    HDISIPOVTD DKTWSSHNFL PSNGVCPQPK TFHVFGRQYR ASYEPLCVFA
50
                    EKIRFAVLLA FIIMSAFVVF GSLGGE*
               501
```

After further analysis, the following gonococcal DNA sequence <SEQ ID 275> was identified:

	1	<b>ATGGTCACAA</b>	AACATACAAA	TTTGAATTTT	GCGAAATTGT	CGATAATTGC
	51	<b>AATTTTGATG</b>	ATGTATTCGT	TTGAAGCGAA	TGCAAATGCA	GTAAAAATAT
	101	CTGAAACTCT	TTCGGTTGAT	ACCGGACAAG	GCGCGAAAGT	TCATAAGTTC
55	151	GTTCCTAAAT	CAAGTAATAT	TTATTCATCT	GATTTAACAA	AAGCGGTAGA
	201	TTTAACGCAT	<b>ATCCCCACGG</b>	GCGCAAAAGC	CCGAATCAAC	GCCAAAATAA
	251	CCGCCAGCGT	ATCCCGCGCC	GGCGTATTGT	CGGGGGTCGG	CAAACTTGTC
	301	CGCCAAGGCG	CGAAATTCGG	CACAAGGGCG	GTTCCCTATG	TCGGAACAGC
	351	CCTTTTAGCC	CACGACGTAT	ACGAAACTTT	CAAAGAAGAC	ATACAGGCAC
60	401	GAGGCTGCCG	<b>ATACGATCCC</b>	GAAACCGACA	AATTT	

This corresponds to the amino acid sequence <SEQ ID 276; ORF72ng-1>:

- 1 MVTKHTNLNF AKLSIIAILM MYSFEANANA VKISETLSVD TGQGAKVHKF 51 VPKSSNIYSS DLTKAVDLTH IPTGAKARIN AKITASVSRA GVLSGVGKLV
- 51 VPKSSNIYSS DLTKAVDLTH 1PTGAKARIN ARTTASVSKA GVESSV 101 RQGAKFGTRA VPYVGTALLA HDVYETFKED 1QARGCRYDP ETDKF
- 5 ORF72ng-1 and ORF721-1 show 89.7% identity in 145 aa overlap:

```
60
                                                       50
                                        30
                        10
        orf72ng-1.pe MVTKHTNLNFAKLSIIAILMMYSFEANANAVKISETLSVDTGQGAKVHKFVPKSSNIYSS
                  MVIKYTNLNFAKLSIIAILMMYSFEANANAVKISETVSVDTGQGAKIHKFVPKNSKTYSS
        orf72-1
                                               40
                                                       50
                                        30
                                20
10
                                                      110
                                               100
                                        90
                        70
                                80
        orf72ng-1.pe DLTKAVDLTHIPTGAKARINAKITASVSRAGVLSGVGKLVRQGAKFGTRAVPYVGTALLA
                  DLIKTVDLTHIPTGAKARINAKITASVSRAGVLAGVGKLARLGAKFSTRAVPYVGTALLA
        orf72-1
15
                                               100
                                                      110
                                        90
                                80
                        130
        orf72ng-1.pe HDVYETFKEDIQARGCRYDPETDKF
                   20
                   HDVYETFKEDIQARGYQYDPETDKFAKVSGX
        orf72-1
                                140
                        130
```

Based on this analysis, including the presence of a putative leader sequence and transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 33

25

The following partial DNA sequence was identified in N.meningitidis <SEQ ID 277>:

```
1 ATGAGATTTT TCGGTATCGG TTTTTTGGTG CTGCTGTTTT TGGAGATTAT
51 GTCGATTGTG TGGGTTGCCG ATTGGCTGGG CGGCGGCTGG ACGTTGTTTT
101 TGATGGCGGC AGGTTTTGCC GCCGGCGTGC TGATGCTCAG GCAAACCGGG
151 GCTGACCGGT CTTTTATTGG CGGCCGCGC AATGAGAAGC GGCGGAAGG
201 TATCCGTTTA TCAGATGTTG TGGCCTATC..
```

- 35 This corresponds to the amino acid sequence <SEQ ID 278; ORF73>:
  - 1 MRFFGIGFLV LLFLEIMSIV WVADWLGGGW TLFLMAAGFA AGVLMLRQTG

51 LTGLLLAGAA MRSGGKVSVY QMLWPI..

Further work revealed the complete nucleotide sequence <SEQ ID 279>:

	1	ATGAGATTTT	TCGGTATCGG	TTTTTTGGTG	CTGCTGTTTT	TGGAGATTAT
40	51	CTCCATTCTC	TEGETTECCE	ATTGGCTGGG	CGGCGGCTGG	ACGTTGTTTT
40	101	TCATCCCCCC	AGGTTTTGCC	GCCGGCGTGC	TGATGCTCAG	GCATACGGGG
	151	CTGTCCGGTC	TTTTATTGGC	GGGCGCGGCA	ATGAGAAGCG	GCGGGAGGGT
	201	ATCCGTTTAT	CAGATGTTGT	GGCCTATCCG	TTATACGGTG	ATTICCTICT CTG1
	251	GTCTGATGAG	TCCGGGATTC	GTATCCTCGG	CCACCACGTG	CCCDDDATTT
45	301	CTGCCGTTTA	AGGGAGGGGC	AGTGTTGCAG GCAGAAAAGA	GCAGGAGGIG	CGCGATGACG
	351	TTTCAACATG	AACCAATCGG	ACGGTTGAAG	AGCCTTACGG	CGGCAATCGT
	401	ATATTATCGA	CCATCCAACA	CAAAAAAAGAC	GAATAA	
	451	TCCCGAAACG	CCALCGMACM			

This corresponds to the amino acid sequence <SEQ ID 280; ORF73-1>:

50 1 MRFFGIGFLV LLFLEIMSIV WVADWLGGGW TLFLMAAGFA AGVLMLRHTG
51 LSGLLLAGAA MRSGGRVSVY QMLWPIRYTV AAVCLMSPGF VSSVLAVLLL
101 LPFKGGAVLQ AGGAENFFNM NQSGRKEGFS RDDDIIEGEY TVEEPYGGNR

#### 151 SRNAIEHKKD E\*

Computer analysis of this amino acid sequence gave the following results:

### Homology with a predicted ORF from N. meningitidis (strain A)

ORF73 shows 90.8% identity over a 76aa overlap with an ORF (ORF73a) from strain A of N.

#### 5 meningitidis:

30

		10	20	30	40	50	60
	orf73.pep	MRFFGIGFLVLLFLEI	MSIVW	/ADWLGGGWTLFL	MAAGFAAGVI	MLRQTGLTGL	LLAGAA
	• •	111111111111111	111111		п. нн:	111:111:11	
	orf73a	MRFFGIGFLVLLFLEI	MSIVW	/ADWLGGGWTLFL	<u>Maatfa</u> agvv	MLRHTGLSGL	LLAGAA
10		10	20	30	40	50	60
		70					
	orf73.pep	MRSGGKVSVYQMLWPI					
		11111:3111 111 1					
15	orf73a	MRSGGRVSVYXMLWXI	RYTVA	AVC <u>XMSPGFVSSV</u>	XAVLLXLPF1	KGGAVLQAGGA	ENFFNM
		•					

#### The complete length ORF73a nucleotide sequence <SEQ ID 281> is:

	1	ATGAGATTTT	TCGGTATCGG	TTTTTTGGTG	CTGCTGTTTT	TGGAGATTAT
	51	GTCGATTGTG	TGGGTTGCCG	ATTGGTTGGG	CGGCGGTTGG	ACGCTGTTTC
	101	TAATGGCGGC	AACCTTTGCC	GCCGGCGTGG	TGATGCTCAG	GCATACGGGG
20	151	CTGTCCGGTC	TTTTATTGGC	GGGCGCGGCA	ATGAGAAGCG	GCGGGAGGGT
	201	ATCCGTTTAT	CANATGTTGT	GGCNTATCCG	TTATACGGTG	GCGGCGGTGT
	251	GTCNGATGAG	TCCGGGATTC	GTATCCTCGG	TGTNGGCGGT	ATTGCTGNTG
	301	CTNCCGTTTA	AGGGAGGTGC	AGTGTTGCAG	GCAGGAGGTG	CGGAAAATTT
	351	TTTCAACATG	AACCANTCGG	GCAGAAAAGA	NGGCNTTTCC	CGCGATGACG
25	401	ATATTATCGA	GGGGGAATAT	ACGGTTGAAG	ANCCTTACGG	CGGCANTCGT
	451	TTCCGAAACG	CCNTNGAACA	CAAAAAAGAC	GAATAA	

This encodes a protein having amino acid sequence <SEQ ID 282>:

1 MRFFGIGFLV LLFLEIMSIV WVADWLGGGW TLFLMAATFA AGVVMLRHTG
51 LSGLLLAGAA MRSGGRVSVY XMLWXIRYTV AAVCXMSPGF VSSVXAVLLX
101 LPFKGGAVLQ AGGAENFFNM NXSGRKXGXS RDDDIIEGEY TVEXPYGGXR
151 FRNAXEHKKD E\*

### ORF73a and ORF73-1 show 91.3% identity in 161 aa overlap

35	orf73a.pep			30 ADWLGGGWTLFI           ADWLGGGWTLFI	HIL     : MAAGFAAGVI	:             LMLRHTGLSGL	 LLAGAA
		10	20	30	40	50	60
		70	80	90 .	100	110	120
40	orf73a.pep	MRSGGRVSVYX	MLWXIRYTVAA	VCXMSPGFVSS	/XAVLLXLPF1	KGGAVLQAGGA	
		111111111					
	orf73-1			VCLMSPGFVSS <sup>1</sup>			
		70	80	90	100	110	120
45		130	140	150	160		
	orf73a.pep	NXSGRKXGXSR	DDDIIEGEYTV	EXPYGGXRFRN	AXEHKKDEX		
		1 1111 1 11		1 1111 1 11	1 1111111		
	orf73-1	NOSGRKEGFSR	DDDIIEGEYTV	EEPYGGNRSRN	AIEHKKDEX		
		130	140	150·	160		
50							

### Homology with a predicted ORF from N. gonorrhoeae

ORF73 shows 92.1% identity over a 76aa overlap with a predicted ORF (ORF73.ng) from N. gonorrhoeae:

orf73.pep MRFFGIGFLVLLFLEIMSIVWVADWLGGGWTLFLMAAGFAAGVLMLRQTGLTGLLLAGAA 60

50

55

	orf73ng MRFFGIGFLVLLFLEIMSIVWVADWLGGGWTLFLMAATFAAGVLMLRHTGLSGLLLAGAA 60	
	orf73.pep MRSGGKVSVYQMLWPI 76	
5	:: :	
3	The complete length ORF73ng nucleotide sequence <seq 283="" id=""> is:</seq>	
	A PROPORTING TOCCOPATIONS THEFTTOGGTG CTGCTGTTTT TGGAAATTAT	
	51 GTCGATTGTG TGGGTTGCCG ATTGGCTGGG CGGCGGTTGG ACGCTGTTCC	
10	and amanagacine memeratence transfer GTAAAAagta ququanaggi	
	201 ATCTGTTTAT CagatgtTGT GGCCTATCCG TTATAcggtg gcggcggtgT 251 GTCTGatgag tCcggGATTC GTATCCTccg tgttggCGGT ATTGCTGCTG	
	301 CTGCcgttta aggGaggGgc agtgttgcag gcaggaggtg cggadanii	
15	And nemetations duramagnatiat according and accordacy cycladicy c	
13	451 tcccgaAAcg ccatcgaaca cgaaaAagac gaataA	
	This encodes a protein having amino acid sequence <seq 284="" id="">:</seq>	
	1 MRFFGIGFLV LLFLEIMSIV WVADWLGGGW TLFLMAATFA AGVLMLRHTG 51 LSGLLLAGAA VKSSGKVSVY QMLWPIRYTV AAVCLMSPGF VSSVLAVLLL	
20	101 LPFKGGAVLQ AGGAENFFNM NQSGRKEGFF HDDDIIEGEY TVERPDGGNR	
20	151 SRNAIEHEKD E*	
	ORF73ng and ORG73-1 show 93.8% identity in 161 aa overlap	
	10 20 30 40 50 60 orf73-1.pep MRFFGIGFLVLLFLEIMSIVWVADWLGGGWTLFLMAAGFAAGVLMLRHTGLSGLLLAGAA	
25		
	orf73ng MRFFGIGFLVLLFLEIMSIVWVADWLGGGWTLFLMAATFAAGVLMLRHTGLSGLLLAGAA 10 20 30 40 50 60	
	70 80 90 100 110 120	
30	orf73-1.pep MRSGGRVSVYQMLWPIRYTVAAVCLMSPGFVSSVLAVLLLLPFKGGAVLQAGGAENFFNM ::::::::::::::::::::::::::::::::::	
	orf73ng VKSSGKVSVYQMLWPIRYTVAAVCLMSPGFVSSVLAVLLLLPFRGGAVLAGAGALETT IN 120	
	70 00 100	
35	NOSGRKEGFSRDDDIIEGEYTVEEPYGGNRSRNAIEHKKDEX	
	130 140 150 160	
40	Based on this analysis, including the presence of a putative leader sequence and putativ	е
	transmembrane domain in the gonococcal protein, it is predicted that the proteins from	n
	N.meningitidis and N.gonorrhoeae, and their epitopes, could be useful antigens for vaccines of	r
	diagnostics, or for raising antibodies.	
	<u></u>	
	Example 34	
45		
	1 ATGTTTGTTT TTCAGACGGC ATTCTT.ATG TTTCAGAAAC ATTTTGCAGAA	
	ATGITTGTT TICAMOUNT TO AGGERACATT ATACGTGGTT GCCACGCCCA  AGCCTCCGAC AGCGTCGTCG GAGGGACATT ATACGTGGTT GCCACGCCCA  TCGGCAATTT GGCGGACATT ACCCTGCGCG CTTTTGGCGGT ATTGCAAAAG	
	151 GCGGCCGA AGACACGCGC GTTACCGCAC AGCTTTTGAG	

451 GATTTTATT TCAACGGTTT TGTACCGCCG AAATCGGGAG AACGCAGGAA

orf75.pep

5	551 601 651 701 751 801 851	ACTGTTTGCC AAATGGGTGC GGGCGGCGTT TCCTATCGTC ATGTTTGAAA CGCCGCACCG CATCGGTGCA GCGCTTGCCG ATATGGCGGA ACTGTTCCCC GAACGCCGAT TAATGCTGGC GCGCGAAATT ACGAAAACGT TTGAAACGTT CTTAAGCGGC ACGGTTGGGG AAATTCAGAC GGCATTGTCT GCCGACGGCG ACCAATCGCG CGGCGAGATG GTGTTGGTGC TTTATCCGGC GCAGGATGAA AAACACGAAG GCTTGTCCGA GTCCGCGCAA AACATCATGA AAATCCTCAC AGCCGAGCTG CCGACCAAAC AGGCGGCGGA GCTTGCTGCC AAAATCACGG GCGAGGGAAA GAAAGCTTTG TACGAT.
	This corresponds	s to the amino acid sequence <seq 286;="" id="" orf75="">:</seq>
10 15	51 101 151 201	MFVFQTAFXM FQKHLQKASD SVVGGTLYVV ATPIGNLADI TLRALAVLQK AAEDTR VTAQLLSAYG IQGKLVSVRE HNERQMADKI VGYLSDGMVV AQVSDAGTPA VCDPGAKLAR RVREAGFKVV PVVGAXAVMA ALSVAGVEGS DFYFNGFVPP KSGERRKLFA KWVRAAFPIV MFETPHRIGA ALADMAELFP ERRLMLAREI TKTFETFLSG TVGEIQTALS ADGDQSRGEM VLVLYPAQDE KHEGLSESAQ NIMKILTAEL PTKQAAELAA KITGEGKKAL YD
		realed the complete nucleotide sequence <seq 287="" id="">:</seq>
	1 51	ATGTTTCAGA AACATTTGCA GAAAGCCTCC GACAGCGTCG TCGGAGGGAC ATTATACGTG GTTGCCACGC CCATCGGCAA TTTGGCGGAC ATTACCCTGC
20	101 151 201	GCGCTTTGGC GGTATTGCAA AAGGCGGACA TCATCTGTGC CGAAGACACG CGCGTTACCG CACAGCTTTT GAGCGCGTAC GGCATTCAGG GCAAACTCGT CAGTGTGCGC GAACACAACG AACGGCAGAT GGCGGACAAG ATTGTCGGCT
25	301 351	GTTTAAAGTC GTTCCCGTCG TGGGCGCAAG CGCGGTGATG GCGGCTTTGA
25	401 451 501 551	CCGAAATCGG GAGAACGCAG GAAACTGTTT GCCAAATGGG TGCGGGCGGC GTTTCCTATC GTCATGTTTG AAACGCCGCA CCGCATCGGT GCGACGCTTG
30	601 651 701	ATTACGAAAA CGTTTGAAAC GTTCTTAAGC GGCACGGTTG GGGAAATTCA GACGGCATTG TCTGCCGACG GCAACCAATC GCGCGGCGAG ATGGTGTTGG TGCTTTATCC GGCGCAGGAT GAAAAACACG AAGGCTTGTC CGAGTCCGCG
	751 801 851	CAAAACATCA TGAAAATCCT CACAGCCGAG CTGCCGACCA AACAGGCGGC GGAGCTTGCT GCCAAAATCA CGGGCGAGGG AAAGAAAGCT TTGTACGATC TGGCTCTGTC TTGGAAAAAC AAATAG
35	This corresponds	s to the amino acid sequence <seq 288;="" id="" orf75-1="">:</seq>
40	51 101 151 201	ITKTFETFLS GTVGEIQTAL SADGNQSRGE MVLVLYPAQD EKHEGLSESA
	251 Computer analys	QNIMKILTAE LPTKQAAELA AKITGEGKKA LYDLALSWKN K* sis of this amino acid sequence gave the following results:
	Homology with	a predicted ORF from N.meningitidis (strain A)
	ORF75 shows 9	5.8% identity over a 283aa overlap with an ORF (ORF75a) from strain A of N.
45	meningitidis:	
	orf75.pep	10 20 30 40 50 60 MFVFQTAFXMFQKHLQKASDSVVGGTLYVVATPIGNLADITLRALAVLQKAXXXXAEDTR
50	orf75a	MFQKHLQKASDSVVGGTLYVVATPIGNLADITLRALAVLQKADIICAEDTR 10 20 30 40 50
<i>.</i>	orf75.pep	
55	orf75a	VTAQLLSAYGIQGKLVSVREHNERQMADKIVGYLSDGMVVAQVSDAGTPAVCDPGAKLAR 60 70 80 90 100 110

130 140 150 160 170 180 RVREAGFK<u>VVPVVGAXAVMAALSVA</u>GVEGSDFYFNGFVPPKSGERRKLFAKWVRAAFPIV

PCT/IB98/01665 WO 99/24578

-198-

								:111:1
	orf75a	:    RVREVGFK	VVPVVGASAV	MAALSVA	VAGSDEYE	NGLALLVOGE	KKKK TE LIKA A L	VAFPVV 170
	011734	12		10	140	150	160	1,0
5		1	90 2	200	210	220	230 TOTALSADGE	240 OSRGEM
J	orf75.pep						IOTALSADGE	11111
	orf75a	MFETPHRI	GATLADMAEI	JFPERRLM.	LAREITKTE 200	TETFLSGTVGE 210	IQTALAADGN 220	IQSRGEM 230
10		18	30 13	90				
10		2	250	260 RAONTMKT	270 t.taet.ptko	280 DAAELAAKIT	290 SEGKKALYD	
	orf75.pep							AT CHIKNIK
	orf75a	VLVLYPA(	DEKHEGLSE:	SAQNIMKI 50	LTAELPTK( 260	QAAELAAKIT 270	SEGKKALYDL 280	290
15		2.	40 Z	<b>J</b> 0				
	orf75a	x				'		
	The complete len	ngth ORF75	a nucleotid	e sequen	ce <seq< th=""><th>ID 289&gt; is:</th><th></th><th></th></seq<>	ID 289> is:		
	_		አ አ <b>ሮ አ ም</b> ሞምር ር	A GAAAGO	CTCC GAC	AGCGTCG TC	GGAGGGAC	
20	1 51		この中へへへるべる	C CCATC	CCAA TTT	GGCGGAC AT	IACCCIGC	
	101 151	GCGCTTTGGC CGCGTTACCG	• ୯୯୯୬ ୯୯୩୩	TH GAGCG	GTAC GGC	ATTUAGG GU	WWWCICGI	
	201	CAGCGTTACCG CAGCGTGCGC ATCTTTCAGA	• ሮአአሮአሮአሽር	יה אארההנ	'AGAT GGC	GGACAAG A1	TGTCGGCT	
0.5	251 301		* xcccccccc	C GAAAC	rcgcc cgc	CGCGTGC G1	GROGICGG	
25	351	GTTTAAAGTT GTGTGGCTGG	, CMCCCMCM4	יכ יירכככנ	GCAAG CGU	GGTGATG G	GGCIIIGN	
	401 451	~~~~ T T T C C C	•	ас сават	rcatr GCC	AAATGGG I	3CGGG1GGC	
	501		- ሮጥርአጥርጥጥ	rc AAACC	CCGCA CCG	CATCEGE G	CACGCIIG	
30	551 601		. CCTTTCAA	ልሮ ርጥጥርጥ	TAAGC GGU	TTAATGCT GCCACGGTTG GC	GHANTICA	
	651		~ ~~~~~~~~~	<b>ሮር ሮሮልልሮ</b>	CAATC GCC	iCGGCGAG A	IGGIGIIGG	
	701 751		• mczzzmc	ሮጥ ሮኳሮልር	CCGAG CTC	GCCTTGTC C	MCMGGGGGG	
35	801	GGAGCTTGC	C GCCAAAAT	CA CGGGC	GAGGG AA	AAAAAGCT T	TGTACGATC	
	851		C TTGGAAAA			Ω ID 200~.		
	This encodes a	protein havi	ng amino a	cid seque	ence <se< th=""><th>Q ID 2902.</th><th></th><th></th></se<>	Q ID 2902.		
	1	MFQKHLQKA	S DSVVGGTL	YV VATPI	GNLAD IT	LRALAVLO K	ADIICAEDT	
	51	RVTAQLLSA	Y GIQGKLVS	VR EHNEF	ROMADK IV Basavm aa	LSVAGVAG S	DFYFNGFVP	
40	101 151	STANDERDER	E VALUATION E	יידא אלו עסי	PHRIG AT	LAUMAELE E	CKKTMTWKE	
	201	ITKTFETFL	S GTVGEIQT	TAL AADGI	NOSRGE MV SEGKKA LY	LVLYPAQD E DLALSWKN F	* 'Vuegesesv	
	251							
	ORF75a and O	)RF75-1 sno	W 98.3% IC	lenniy m	291 aa 0	veriap.		
45			10	20	30	40 rt dat.avt.ok?	50 ADIICAEDTR	60 TAOLLSAY
	orf75a.p							111111
	orf75-1	MFQKH:	LQKASDSVVG	GTLYVVAT	PIGNLADIT 30	rlralavloki 40	ADIICAEDTR'	VTAQLLSAY 60
50			10	20				120
30	•		70	80 0MX DK TVG	90 YT.SDGMVV	100 AOVSDAGTPA	110 VCDPGAKLAR	
	orf75a.				1 1 1 1 1 1 1 1	1111111111		111111
	orf75-1		LVSVREHNER 70	QMADKIVO 80	YLSDGMVV 90	AQVSDAGTPA 100	VCDPGAKLAR 110	120
55			70	-			170	180
		TOUR	130	140 VAGVAGSDE	150 TYFNGFVPP	160 KSGERRKLFA	KWVRVAFPVV	MFETPHRIG
	orf75a.	11111				11111111111	11111111	11111111
60	orf75-1	VPVVG	ASAVMAALSV 130	/AGVEGSDI 140	FYFNGEVPP 150	KSGEKKKLFF 160	KWVRAAFPIV 170	180
				200	210	220	230	240
		pep m ATLAI	190 MAELFPERRI	200 LMLAREIT	หากการการเก	TVGEIOTAL	ADGNOSRGEN	VLVLYPAQD
65			ппппп	11111111	111111111		:11111111111	
, , ,								

		AAAcgca cgctcgaATT					
		acGCGCc gactaCAATC					
		SCACCAC GCCCAAacTT					
5		STGGCGC ATCTCGCTGA					
J		GCCGCT TTCCTATTTC	AACCCGCGCA	GCGGACATAC	CTACAATATC	, 2	
		ATTGCCA TCGGTTTGTT					
		TGAAGCC GTGGAAGACG					
• •		IGCACAT CATCATGTTC					
10		ATGCCCA GCCAGCCCTT	CTGGCAGGCG	GTTGGCAAAA	GTCTGACATT	:	
	1101 GAA	AGgcgGA AAATGA					
	This corresponds to t	he amino acid seque	nce <seo i<="" th=""><th>D 506; ORF</th><th>101ng-1&gt;:</th><th></th><th></th></seo>	D 506; ORF	101ng-1>:		
		•		,			
	1 MIY	ORNLIKE LSFTAVGIFV	VLLAVLVSTQ	<u>A</u> INLLGRAAD	GRVAIDAVLA	<u>1</u>	
		FWVIGMT PLLLVLTAFI					
15	101 PVM	OFAVPFA ILIAVMOLWV	<u>I</u> PWAELRSRE	YAEILKQKQE	LSLVEAGEFN	1	
	151 NLG 201 NKR'	KRNGRVY FVETFDTESG FLELRHG YRYSGTPGRA	TMKNTLTKFÖ	DKNGGDNIIF	AKEGNESLKI	) r	
		OLIGSSN POHOAELMWR					
•		IGLFLIY QNGLTLLFEA	VEDGKIHFWL	GLLPMHIIMF	VIAIVLLRV	₹	
20		SQPFWQA VGKSLTLKGG					
	ORF101ng-1 and OF	E101 1 chow 07 60	(identity in	271 on overel	O		
	OKTIVING-1 and Or	C 101-1 SHOW 97.07	o identity in	3/1 aa overi	ap:		
		10	20	30	40 5	50 !	60
	orf101-1.pep	MIYQRNLIKELSFTAV					
				HIHHHH	1111111111	11111111	PH1
25	orf101ng-1	MIYQRNLIKELSFTAV	GIFVVLLAVLV	STQAINLLGRA	ADGRVAID <b>A</b> VI	LALVGFWVI	GMT
		. 10	20	30	40 5	50	60
		70	80	90 · 1	00 1	10	120
•	orf101-1.pep	PLLLVLTAFISTLTVL				FAVLVAVMO	
30			11111111111	11111111111	1111111111	11:1:111	HI
	orf101ng-1	PLLLVLTAFISTLTVL					
		70	80	90 1	00 1:	10	120
		130	140	150 1	60 1	70	180
35	orf101-1.pep	IPWAELRSREYAEILK	QKQELSLVEAG	EFNSLGKRNGR	VYFVETFDTE	SGIMKNLFL	REQ
				111:1111111			Ш
	orf101ng-1	IPWAELRSREYAEILK					
		130	140	150 1	60 1	70	180
40		190	200	210 2	20 2:	30	240
	orf101-1.pep			RHGYRYSGTPG	RADYNQVSFQI	KLNLIISTT	PKL -
		- {					
	orf101ng-1	DKNGGDNIIFAKEGNF					
45		190	200	210 2	20 2:	30	240
15		250	260	270 2	80 2	90	300
•	orf101-1.pep	IDPVSHRRTIPTAQLI	GSSNPQHQAEL				
		- ининин ийи					
50	orf101ng-1	IDPVSHRRTISTAQLI					
30		250	260	270 2	80 2	90	300
		310	320	330 3	40 3	50	360
	orf101-1.pep	LIAIGLFLIYQNGLTL	LFEAVEDGKIH	FWLGLLPMHII			
55	orf101ng-1	LIAIGLFLIYQNGLTL					
		310	320	330 3	40 3	50	360
		370					
	orf101-1.pep	VGKSLTLKGGKX					
60		111111111111					
	orf101ng-1	VGKSLTLKGGKX			•	•	
		370					
	Danad on Alvin amplem	ia inaludina tha pro				7.1 1.1	1 1

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is

predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 60

50

The following partial DNA sequence was identified in N.meningitidis <SEQ ID 507>:

```
..GGTGGTGGTT TTATCAATGC TTCCTGTGCC ACTTTGACGA CAGCCAAACC
5
                       GCAATATCAA GCAGGAGACC TTAGCGCTTT TAAGATAAGG CAAGGCAATG
                 51
                       TTGTAATCGC CGGACACGGT TTGGATGCAC GTGATACCGA TTACACACGT
                101
                       ATTCTCAGTT ATCATTCCAA AATCGATGCA CCCGTATGGG GACAAGATGT
                151
                       TCGTGTCGTC GCGGGACAAA ACGATGTGGC CGCAACAGGT GATGCACATT
                201
                       CGCCTATTCT CAATAATGCT GCTGCCAATA CGTCAAACAA TACAGCCAAC
10
                251
                       AACGGCACAC ATATCCCTTT ATTTGCGATT GATACAGGCA AATTAGGAGG
                301
                       TAT.GTATGC CAACAAAATC ACCTTGATCA GTACGGTCGA GCAAGCAGGC
                351
                       ATTCGTAA
                401
```

This corresponds to the amino acid sequence <SEQ ID 508; ORF113>:

```
1 ..GGGFINASCA TLTTAKPQYQ AGDLSAFKIR QGNVVIAGHG LDARDTDYTR
51 ILSYHSKIDA PVWGQDVRVV AGQNDVAATG DAHSPILNNA AANTSNNTAN
101 NGTHIPLFAI DTGKLGGXVC QQNHLDQYGR ASRHS*
```

Computer analysis of this amino acid sequence gave the following results:

Homology with with pspA putative secreted protein of N. meningitidis (accession AF030941)

20 ORF and pspA show 44% aa identity in 179aa overlap:

```
GGGFINASCATLTTAKPQYQAGDLSAFKIRQGNVVIAGHGLDARDTDYTRILSYHSKIDA 60
                                       G+L+ F + G VVI G GLD D DYTRILS ++I+A
                   GGG INA+ TLT+ P
                   GGGLINAASVTLTSGVPVLNNGNLTGFDVSSGKVVIGGKGLDTSDADYTRILSRAAEINA 256
          pspa
                   PVWGQDVRVVAGQNDVAATGDAHSPILXXXXXXXXXXXXXXGTHIPLFAIDTGKLGGMYA 120
25
          orf113
                                                            + P AIDT LGGMYA
                    VWG+DV+VV+G+N +
                   GVWGKDVKVVSGKNKLDFDG-----SLAKTASAPSSSDSVTPTVAIDTATLGGMYA 307
          pspa
                   NKITLISTVEQAGIRNQGQWFASAGNVAVNAEGKLVNTGMIAATGENHAVSLHARNVHN 179
          orf113
                                                                   +++ A+ V N
                             A IRN+G+ FA+ G V ++A+GKL N+G I A
                   +KITLIST
30
                   DKITLISTDNGAVIRNKGRIFAATGGVTLSADGKLSNSGSIDAA----EITISAQTVDN 362
          pspa
```

## Homology with a predicted ORF from N.gonorrhoeae

ORF113 shows 86.5% identity in 52aa overlap at the N- terminal part and 94.1% identity in 17aa

overlap at the C-terminal part with a predicted ORF (ORF113ng) from N. gonorrhoeae:

	orf113	GGGFINASCATLTTAKPQYQAGDLSAFKIR			
	orf113ng	SHPSQLNGYIEVGGRRAEVVIANPAGIAVNGGGFINASRATLTTGQPQYQAGDFSGFKIR	224		
40	orf113	QGNVVIAGHGLDARDTDYTRILSYHSKIDAPVWGQDVRVVAGQNDVAATGDAHSPILNNA	90		
	orf113ng	QGNAVIAGHGLDARDTDFTRILVCQQNHLDQYGRTSRHS 263			
45	orf113	IDTGKLGGXVCQQNHLDQYGRASRHS 135			
45	orf113ng	DFSGFKIRQGNAVIAGHGLDARDTDFTRILVCQQNHLDQYGRTSRHS 263			

The complete length ORF113ng nucleotide sequence <SEQ ID 509> is predicted to encode a protein having amino acid sequence <SEQ ID 510>:

<sup>1</sup> MNKTLYRVIF NRKRGAVVAV AETTKREGKS CADSGSGSVY VKSVSFIPTH 51 SKAFCFSALG FSLCLALGTV NIAFADGIIT DKAAPKTQQA TILQTGNGIP

- 101 QVNIQTPTSA GVSVNQYAQF DVGNRGAILN NSRSNTQTQL GGWIQGNPWL
- 151 TRGEARVVVN QINSSHPSQL NGYIEVGGRR AEVVIANPAG IAVNGGGFIN
- 201 ASRATLTTGQ PQYQAGDFSG FKIRQGNAVI AGHGLDARDT DFTRILVCQQ
- 251 NHLDQYGRTS RHS\*
- Based on this analysis, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 61

The following partial DNA sequence was identified in N.meningitidis <SEQ ID 511>:

```
..TCAACGGGAC ATAGCGAACA AAATTACACT TTGCCGCGAG AAATCACACG
10
                51
                      101
                      GCCATCATGC GCCCAGCCAA GGCACTGAGT TGCCGCAAAG CAACGGTATT
               151
                      TCGCTACCCT ATACGTCCAA TTCTTTTACC CCATTACCCA GCAGCAGCTT
               201
                      ATACATTATC AATCCTGTCA ATAAAGGCTA TCTTGTTGAA ACCGATCCAC
               251
                      GCTTTGCCAA CTACCGTCAA TGGTTGGGTA GTGACTATAT GCLGGACAGC
15
               301
                      CTCAAACTAG ACCCAAACAA TTTACATAAA CGTTTGGGTG ATGGTTATTA
                      CGAGCAACGT TTAATCAATG AACAAATCGC AGAGCTGACA GGGCATCGTC
               351
                      GTTTAGACGG TTATCAAAAC GACGAAGAAC AATTTAAAGC CTTAATGGAT
               401
               451
                      AATGGCGCGA CTGCGGCACG TTcGATGAAT CTCAGCGTTG GCATTGCATT
                      AAGTGCCGAG CAAGTAGCGC AACTGACCAG CGATATTGTT TGGTTGGTAC
               501
20
               551
                      AAAAAGAAGT TAAGCTTCCT GATGGCGGCA CACAAACCGT ATTGGTGCCA
               601
                      CAGGTTTATG TACGCGTTAA AAATGGCGAC ATAGACGGTA AAGGTGCATT
               651
                      GTTGTCAGGC AGCAATACAC AAATCAATGT TTCAGGCAGC CTGAAAAACT
                      CAGGCACGAT TGCAGGGCGC AATGCGCTTA TTATCAATAC CGATACGCTA
               701
               751
                      GACAATATCG GTGGGCGTAT TCATGCGCAA AAATCAGCGG TTACGGCCAC
25
               801
                      ACAAGACATC AATAATATTG GCGGCATGCT TTCTGCCGAA CAGACATTAT
                      TGCTCAACGC AGGCAACAAC ATCAACAGCC AAAGCACCAC CGCCAGCAGT
               851
               901
                      CAAAATACAC AAGGCAGCAG CACCTACCTA GACCGAATGG CAGGTATTTA
               951
                      TATCACAGGC AAAGAAAAAG GTGTTT...
```

This corresponds to the amino acid sequence <SEQ ID 512; ORF115>:

```
1 ..stghseqnyt Lpreitrnis Lgsfayeshr kalshhapsq gtelpqsngi
51 slpytsnsft plpssslyii npvnkgylve tdprfanyrq wlgsdymlds
101 lkldpnnlhk rlgdgyyeqr lineqiaelt ghrrldgyqn deeqfkalmd
151 ngataarsmn lsvgialsae qvaqltsdiv wlvqkevklp dggtqtvlvp
201 qvyvrvkngd ldgkgallsg sntqinvsgs lknsgtiagr naliintdtl
251 dniggrihaq ksavtatqdi nniggmlsae qtllnagnn insqsttass
301 qntqgsstyl drmagiyitg kekgv..
```

Computer analysis of this amino acid sequence gave the following results:

Homology with the pspA putative secreted protein of *N.meningitidis* (accession number AF030941) ORF115 and pspA protein show 50% aa identity in 325aa overlap:

40	Orf115:	1	STGHSEQNYTLPREITRNISLGSFAYESHRKALSHHAPSQGTELPQSNGISLPYTSNSFT 6 STG+S Y E++ +I +G AY+ + + P + NGI +T	0
	pspA:	778	STGYSRSPYEPAPEVS-SIRMGISAYKGYAPQQASDIPGTVVPVVAENGIHPTFT 8	31
45	Orf115:	61	PLPSSSLYIINPVNKGYLVETDPRFANYRQWLGSDYMLDSLKLDPNNLHKRLGDGYYEQR 1	.20
43	pspA:	832	LP+SSL+ I P NKGYL+ETDP F +YR+WLGS YML +L+ DPN++HKRLGDGYYEQ+ -LPNSSLFAIAPNNKGYLIETDPAFTDYRKWLGSGYMLAALQQDPNHIHKRLGDGYYEQK 8	90
	Orf115:	121	LINEQIAELTGHRRLDGYQNDEEQFKALMDNGATAARSMNLSVGIALSAEQVAQLTSDIV 1	.80
50	pspA:	891	L+NEQIA+LTG+RRLDGY NDEEQFKALMDNG T A+ + L+ GIALSAEQVA+LTSDIV LVNEQIAKLTGYRRLDGYTNDEEQFKALMDNGITIAKELQLTPGIALSAEQVARLTSDIV 9	50
·	Orf115:	181	WLVQKEVKLPDGGTQTVLVPQVYVRVKNGDIDGKGALLSGSNTQINVSGSLKN-SGTIAG 2	39
F. F.	pspA:	951	WL + V LPDG TQTVL P+VYVR + D++G+GALLSGS I SG+++N G IAG WLENETVTLPDGTTQTVLKPKVYVRARPKDMNGQGALLSGSVVDIG-SGAIENRGGLIAG 1	.009
55			·	

```
RNALIINTDTLDNIGGRIHAQKSAVTATQDINNIGGMLSAEQTLLLNAGXXXXXXXXX 299
          Orf115: 240
                                              A DING + AE LLL A
                      R ALI+N
                               + N+ G + +
                 1010 REALILNAQNIKNLQGDLQGKNIFAAAGSDITNTGS-IGAENALLLKASNNIESRSETRS 1068
          pspA:
5
          Olf115: 300 XXXXXXXXXYLDRMAGIYITGKEKG 324
                                + R+AGIY+TG++ G
                 1069 NQNEQGSVRNIGRVAGIYLTGRQNG 1093
          pspA:
     Homology with a predicted ORF from N.gonorrhoeae
     ORF115 shows 91.9% identity over a 334aa overlap with a predicted ORF (ORF115ng) from
10
     N.gonorrhoeae:
                                                                                 31
          orf115.pep
                                                 STGHSEQNYTLPREITRNISLGSFAYESHRK
                                                   111 1111111:1111:11111111111111
                      NEQTFGEKKVFSENGKLHNYWRARRKGHDETGHREQNYTLPEEITRDISLGSFAYESHSK
          orf115ng
                                                                                 71
15
          orf115.pep
                      ALSHHAPSQGTELPQSN-----GISLPYTSNSFTPLPSSSLYIINPVNKGYLVET
                                                                                 81
                       111:11111111111
                                                11,11,11, 11,11,11; 11,11,11; 11,11,11,1
                      ALSRHAPSQGTELPQSNRDNIRTAKSNGISLPYTPNSFTPLPGSSLYIINPANKGYLVET
                                                                                131
          orf115ng
20
                      DPRFANYRQWLGSDYMLDSLKLDPNNLHKRLGDGYYEQRLINEQIAELTGHRRLDGYQND
          orf115.pep
                                                                                141
                       DPRFANYRQWLGSDYMLGSLKLDPNNLHKRLGDGYYEQRLINEQIAELTGHRRLDGYQND
          orfl15ng
                                                                                191
          orf115.pep
                      EEQFKALMDNGATAARSMNLSVGIALSAEQVAQLTSDIVWLVQKEVKLPDGGTQTVLVPQ
                                                                                201
25
                       orf115ng
                      EEQFKALMDNGATAARSMNLSVGIALSAEQAAQLTSDIVWLVQKEVKLPDGGTQTVLMPQ
                                                                                251
          orf115.pep
                      VYVRVKNGDIDGKGALLSGSNTQINVSGSLKNSGTIAGRNALIINTDTLDNIGGRIHAQK
                                                                                261
                       30
          orf115ng
                       VYVRVKNGGIDGKGALLSGSNTQINVSGSLKNSGTIAGRNALIINTDTLDNIGGRIHAOK
          orf115.pep
                       SAVTATQDINNIGGMLSAEQTLLLNAGNNINSQSTTASSQNTQGSSTYLDRMAGIYITGK
                                                                                321
                       SAVTATQDINNIGGILSAEQTLLLNAGNNINNQSTAKSSQNAQGSSTYLDRMAGIYITGK
          orfl15ng
                                                                               371
35
          orf115.pep
                       EKGV
                                                                                 325
                       1111
          orf115ng
                       EKGVLAAQAGKDINIIAGQISNQSDQGQTRLQAGRDINLDTVQTGKYQEIHFDADNHTIR 431
     An ORF115ng nucleotide sequence <SEQ ID 513> was predicted to encode a protein having amino
40
     acid sequence <SEQ ID 514>:
                 1 MLVQTEKDGL HNEQTFGEKK VFSENGKLHN YWRARRKGHD ETGHREQNYT
                51
                    LPEEITRDIS LGSFAYESHS KALSRHAPSQ GTELPQSNRD NIRTAKSNGI
               101
                    SLPYTPNSFT PLPGSSLYII NPANKGYLVE TDPRFANYRQ WLGSDYMLGS
                    LKLDPNNLHK RLGDGYYEQR LINEQIAELT GHRRLDGYQN DEEQFKALMD
               151
45
                   NGATAARSMN LSVGIALSAE QAAQLTSDIV WLVQKEVKLP DGGTQTVLMP
               201
                    QVYVRVKNGG IDGKGALLSG SNTQINVSGS LKNSGTIAGR NALIINTDTL
               251
                    DNIGGRIHAQ KSAVTATQDI NNIGGILSAE QTLLLNAGNN INNQSTAKSS
               301
                    QNAQGSSTYL DRMAGIYITG KEKGVLAAQA GKDINIIAGQ ISNQSDQGQT
               351
                    RLQAGRDINL DTVQTGKYQE IHFDADNHTI RGSTNEVGSS IQTKGDVTLL
               401
 50
                    SGNNLNAKAA EVGSAKGTLA VYAKNDITIS SGIHAGQVDD ASKHTGRSGG
               451
               501
                    GNKLVITDKA QSHHETAQSS TFEGKQVVLQ AGNDANILGS NVISDNGTRI
                    QAGNHVRIGT TOTOSOSETY HOTOKSGLMS AGIGFTIGSK TNTQENOSOS
               551
                    NEHTGSTVGS LKGDTTIVAS KHYEQTGSNV SSPEGNNLIS TQSMDIGAAQ
               601
               651
                    NQLNSKTTQT YEQKGLTVAF SSPVTDLAQQ AIAVAHKAAK QFDKAKTTAL
 55
                    MPWRLPMQVG RLFKQAKAPK K*
               701
      Further work revealed the following partial gonococcal DNA sequence <SEQ ID 515>:
                    TTGCTTGTGC AAACAGAAAA AGACGGTTTG CATAACGAGC AAACCTTTGG
                51
                    CGAGAAGAAA GTCTTCAGCG AAAATGGTAA GTTGCACAAC TACTGGCGTG
                    CGCGTCGTAA AGGACATGAT GAAACAGGGC ATCGTGAACA AAATTATACT
               101
 60
               151
                    TTGCCGGAGG AAATCACACG CGACATTTCA CTGGGTTCAT TTGCCTATGA
```

ATCGCATAGC AAAGCATTAA GCCGTCATGC GCCCAGCCAA GGCACTGAGT

TGCCACAAAG TAACCGGGAT AATATCCGTA CTGCGAAAAG CAACGGTATT

201

```
TCGCTACCCT ATACGCCCAA TTCTTTTACC CCATTACCCG GCAGCAGCTT
               301
               351
                     ATACATTATC AATCCTGCCA ATAAAGGCTA TCTTGTTGAA ACCGATCCAC
               401
                     GCTTTGCCAA CTACCGTCAA TGGTTGGGTA GTGACTATAT GCTGGGCAGC
                     CTCAAACTAG ACCCAAACAA TTTACATAAA CGTTTGGGTG ATGGTTATTA
               451
5
                     CGAGCAACGT TTAATCAATG AACAAATCGC AGAGCTGACA GGGCATCGTC
               501
               551
                     GTTTAGACGG TTATCAAAAC GACGAAGAAC AATTTAAAGC CTTAATGGAT
                601
                     651
                     AAGTGCCGAG CAAGCAGCGC AACTGACCAG CGATATTGTT TGGTTGGTAC
               701
                     AAAAAGAAGT TAAACTTCCT GATGGCGGCA CACAAACCGT ATTGATGCCA
                     CAGGTTTATG TACGCGTTAA AAATGGCGGC ATAGACGGTA AAGGTGCATT
10
               751
               801
                     GTTGTCAGGC AGCAATACAC AAATCAATGT TTCAGGCAGC CTGAAAAACT
                     CAGGCACGAT TGCAGGGCGC AATGCGCTTA TTATCAATAC CGATACGCTA
               851
                901
                     GACAATATCG GTGGGCGTAT TCATGCGCAA AAATCAGCGG TTACGGCCAC
                     ACAAGACATC AATAATATTG GCGGCATTCT TTCTGCCGAA CAGACATTAT
                951
15
               1001
                     TGCTCAATGC GGGTAACAAC ATCAACAACC AAAGCACGGC CAAGAGCAGT
                     CAAAATGCAC AAGGTAGCAG CACCTACCTA GACCGAATGG CAGGTATTTA
               1051
                     TATCACAGGC AAAGAAAAAG GTGTTTTAGC AGCGCAGGCA GGCAAAGACA
               1101
                     TCAACATCAT TGCCGGTCAA ATCAGCAATC AATCAGATCA AGGGCAAACC
               1151
               1201
                     CGGCTGCAGG CAGGACGCGA CATTAACCTG GATACGGTAC AAACCGGCAA
20
                     ATATCAAGAA ATCCATTTTG ATGCCGATAA CCATACCATC CGAGGTTCAA
               1251
               1301
                     CGAACGAAGT CGGCAGCAGC ATTCAAACAA AAGGCGATGT TACCCtatTG
                     TCAGGGAATA ATCTCAATGC CAAAGCTGCC GAAGTCGGCA GCGCAAAAGG
               1351
               1401
                     CACACTTGCC GTGTATGCTA AAAATGACAT TACTATCAGC TCAGGCATCC
                     ATGCCGGCCA AGTTGATGAT GCGTCCAAAC ATACAGGCAG AAGCGGCGGC
               1451
                     GGTAATAAAT TAGTCATTAC CGATAAAGCC CAAAGTCATC ACGAAACTGC
TCAAAGCAGC ACCTTTGAAG GCAAGCAAGT TGTATTGCAG GCAGGAAACG
25
               1501
               1551
               1601
                     ATGCCAACAT CCTTGGCAGT AATGTTATTT CCGATAATGG CACCCGGATT
                     CAAGCAGGCA ATCATGTTCG CATTGGTACA ACCCAAACTC AAAGCCAAAG
               1651
               1701
                     CGAAACCTAT CATCAAACCC AAAAATCAGG ATTGATGAGT GCAGGTATCG
30
                     GCTTCACTAT TGGCAGCAAG ACAAACACAC AAGAAAACCA ATCCCAAAGC
               1751
                     AACGAACATA CAGGCAGTAC CGTAGGCAGC CTGAAAGGCG ATACCACCAT
               1801
                     TGTTGCAAGC AAACACTACG AACAAACCGG CAGCAACGTT TCCAGCCCTG
               1851
               1901
                     AGGGCAACAA CCTTATCAGC ACGCAAAGTA TGGATATTGG CGCAGCACAA
                     AACCAATTAA ACAGCAAAAC CACCCAAACC TACGAACAAA AAGGCTTAAC
               1951
35
               2001
                     GGTGGCATTC AGTTCGCCCG TTACCGATTT GGCACAACAA GCGATTGCCG
                     TAGCACACAA AGCAGCAAAC AAGTCGGACA AAGCAAAAAC GACCGCGTTA
               2051.
               2101
                     ATGCCATGGC GGCTGCCAAT GCAGGTTGGC AGGCCTATCA AACAGGCAAA
               2151
                     GGCGCACAAA ACTTAG
      This corresponds to the amino acid sequence <SEQ ID 516; ORF115ng-1>:
40
                     LLVQTEKDGL HNEQTFGEKK VFSENGKLHN YWRARRKGHD ETGHREONYT
                     LPEEITRDIS LGSFAYESHS KALSRHAPSQ GTELPQSNRD NIRTAKSNGI
                     SLPYTPNSFT PLPGSSLYII NPANKGYLVE TDPRFANYRO WLGSDYMLGS
                     LKLDPNNLHK RLGDGYYEQR LINEQIAELT GHRRLDGYQN DEEQFKALMD
                     NGATAARSMN LSVGIALSAE QAAQLTSDIV WLVQKEVKLP DGGTQTVLMP
QVYVRVKNGG IDGKGALLSG SNTQINVSGS LKNSGTIAGR NALIINTDTL
45
                251
                     DNIGGRIHAQ KSAVTATQDI NNIGGILSAE QTLLLNAGNN INNQSTAKSS
                     QNAQGSSTYL DRMAGIYITG KEKGVLAAQA GKDINIIAGQ ISNQSDQGQT
                351
                     RLQAGRDINL DTVQTGKYQE IHFDADNHTI RGSTNEVGSS IQTKGDVTLL
SGNNLNAKAA EVGSAKGTLA VYAKNDITIS SGIHAGQVDD ASKHTGRSGG
                451
 50
                     GNKLVITDKA QSHHETAQSS TFEGKQVVLQ AGNDANILGS NVISDNGTRI
                     QAGNHVRIGT TOTOSOSETY HOTOKSGLMS AGIGFTIGSK TNTOENOSOS
                551
                601
                     NEHTGSTVGS LKGDTTIVAS KHYEQTGSNV SSPEGNNLIS TQSMDIGAAQ
                     NOLNSKTTOT YEOKGLTVAF SSPVTDLAQO AIAVAHKAAN KSDKAKTTAL
                651
                     MPWRLPMQVG RPIKQAKAHK T*
 55
      This gonococcal protein (ORF115ng-1) shows 91.9% identity with ORF115 over 334aa:
                                          30
                                                              50
           orf115ng-1.p NEQTFGEKKVFSENGKLHNYWRARRKGHDETGHREQNYTLPEEITRDISLGSFAYESHSK
                                                       orfl15
                                                      STGHSEQNYTLPREITRNISLGSFAYESHRK
 60
                                                              10
                                                                         20
                                          90
                                                   100
                                                             110
                                                                        120
           orf115ng-1.p ALSRHAPSQGTELPQSNRDNIRTAKSNGISLPYTPNSFTPLPGSSLYIINPANKGYLVET
                         111:111111111111
                                                   65
           orf115
                         ALSHHAPSQGTELPQSN-
                                                    -GISLPYTSNSFTPLPSSSLYIINPVNKGYLVET
                                                    50
                                                               60
                                                                         70
                                40
                                                                                   80
```

-303-

	orf115ng-1.p	140 150 160 170 180 190 DPRFANYRQWLGSDYMLGSLKLDPNNLHKRLGDGYYEQRLINEQIAELTGHRRLDGYQND	
5	orfl15	DPRFANYRQWLGSDYMLDSLKLDPNNLHKRLGDGYYEQRLINEQIAELTGHRRLDGYQND 90 100 110 120 130 140	
	orfl15ng-1.p	200 210 220 230 240 250 EEQFKALMDNGATAARSMNLSVGIALSAEQAAQLTSDIVWLVQKEVKLPDGGTQTVLMPQ	
10	orf115	EEQFKALMDNGATAARSMNLSVGIALSAEQVAQLTSDIVWLVQKEVKLPDGGTQTVLVPQ 150 160 170 180 190 200	
15	orf115ng-1.p	260 270 280 290 300 310  VYVRVKNGGIDGKGALLSGSNTQINVSGSLKNSGTIAGRNALIINTDTLDNIGGRIHAQK	
20	orf115ng-1.p	320 330 340 350 360 370  SAVTATQDINNIGGILSAEQTLLINAGNNINNQSTAKSSQNAQGSSTYLDRMAGIYITGK	
25	orfl15ng-1.p	380 390 400 410 420 430 EKGVLAAQAGKDINIIAGQISNQSDQGQTRLQAGRDINLDTVQTGKYQEIHFDADNHTIR	
	orf115	EKGV	
	In addition, it show	s homology with a secreted N.meningitidis protein in the database:	
30	= 2273 Score = 60	AF030941) putative secreted protein [Neisseria meningitidis] : 4 bits (1541), Expect = e-172 = 325/678 (47%), Positives = 449/678 (65%), Gaps = 22/678 (3%)	
35		LLVQTEKDGLHNEQTFGEKKVFSENGKLHNYWRARRKGHDETGHREQNYTLPEEITRDIS	
		L+V T + L N++T G K + ++ G LH Y R +KG D TG+ Y E++ I LIVGTPESALDNDETLGTKTI-TDKGDLHRYHRHHKKGRDSTGYSRSPYEPAPEVS-SIR	
40	- •	LGSFAYESHSKALSRHAPSQGTELPQSNRDNIRTAKSNGISLPYTPNSFTPLPGSSLYII +G AY+ + AP Q +++P + NGI +T LP SSL+ I	
		MGISAYKGYAPQQASDIPGTVVPVVAENGIHPTFTLPNSSLFAI NPANKGYLVETDPRFANYRQWLGSDYMLGSLKLDPNNLHKRLGDGYYEQRLINEQIAELT	
45		P NKGYL+ETDP F +YR+WLGS YML +L+ DPN++HKRLGDGYYEQ+L+NEQIA+LT APNNKGYLIETDPAFTDYRKWLGSGYMLAALQQDPNHIHKRLGDGYYEQKLVNEQIAKLT	
	Query: 181	GHRRLDGYQNDEEQFKALMDNGATAARSMNLSVGIALSAEQAAQLTSDIVWLVQKEVKLP	240
50	Sbjct: 901	G+RRLDGY NDEEQFKALMDNG T A+ + L+ GIALSAEQ A+LTSDIVWL + V LP GYRRLDGYTNDEEQFKALMDNGITIAKELQLTPGIALSAEQVARLTSDIVWLENETVTLP	960
50	Query: 241	DGGTQTVLMPQVYVRVKNGGIDGKGALLSGSNTQINVSGSLKN-SGTIAGRNALIINTDT DG TOTVL P+VYVR + ++G+GALLSGS I SG+++N G IAGR ALI+N	299
	Sbjct: 961	DGTTQTVLKPKVYVRARPKDMNGQGALLSGSVVDIG-SGAIENRGGLIAGREALILNAQN	1019
55	_ <del>-</del>	LDNIGGRIHAQKSAVTATQDINNIGGILSAEQTLLLNAGNNINNQSTAKSSQNAQGSSTY + N+ G +	
	-	IKNLQGDLQGKNIFAAAGSDITNTGSI-GAENALLLKASNNIESRSETRSNQNEQGSVRN	
60		$ \begin{array}{llllllllllllllllllllllllllllllllllll$	
	-	IGRVAGIYLTGRQNGSVLLDAGNNIVLTASELTNQSEDGQTVLNAGGDIRSDTTGISRNQ	
	_	EIHFDADNHTIRGSTNEVGSSIQTKGDVTLLSGNNLNAKAAEVGSAKGTLAVYAKNDITI FD+DN+ IR NEVGS+I+T+G+++L + ++ +AAEVGS +G L + A DI +	
65	-	NTIFDSDNYVIRKEQNEVGSTIRTRGNLSLNAKGDIRIRAAEVGSEQGRLKLAAGRDIKV	
	_	SSGIHAGQVDDASKHTGRSGGGNKLVITDKAQSHHETAQSSTFEGKQVVLQAGNDANILG +G + +DA K+TGRSGGG K +T ++ + A S T +GK+++L +G D + G	
70	Sbjct: 1199	EAGKAHTETEDALKYTGRSGGGIKQKMTRHLKNQNGQAVSGTLDGKEIILVSGRDITVTG	1238

```
Query: 540 SNVISDNGTRIQAGNHVRIGTTQTQSQSETYHQTQKSGLM-SAGIGFTIGSKTNTQENQS 598
SN+I+DN T + A N++ + +T+S+S ++ +KSGLM S GIGFT GSK +TQ N+S
Sbjct: 1259 SNIIADNHTILSAKNNIVLKAAETRSRSAEMNKKEKSGLMGSGGIGFTAGSKKDTQTNRS 1318

5 Query: 599 QSNEHTGSTVGSLKGDTTIVASKHYEQTGSNVSSPEGNNLISTQSMDIGAAQNQLNSKTT 658
++ HT S VGSL G+T I A KHY QTGS +SSP+G+ IS+ + I AAQN+ + ++
Sbjct: 1319 ETVSHTESVVGSLNGNTLISAGKHYTQTGSTISSPQGDVGISSGKISIDAAQNRYSQESK 1378

Query: 659 QTYEQKGLTVAFSSPVTD 676
Q YEQKG+TVA S PV +
Sbjct: 1379 QVYEQKGVTVAISVPVVN 1396
```

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 62

35

15 The following partial DNA sequence was identified in N.meningitidis <SEQ ID 517>:

```
..TCAGGGAATA ACCTCAATGC CAAAGCTGCC GAAGTCAGCA GCGCAAACGG
                 51
                       TACACTCGCT GTGTCTGCCA ATAATGACAT CAACATCAGC GCAGGCATCA
                101
                       ACACGACCCA TGTTGATGAT GCGTCCAAAC ACACAGGCAG AAGCGGTGGT
                151
                       GGCAATAAAT TAGTCATTAC CGATAAAGCC CAAAGTCATC ACGAAACCGC
20
                201
                       CCAAAGCAGC ACCTTTGAAG GCAAGCAAGT TGTATTGCAG GCAGGAAACG
                251
                       ATGCCAACAT CCTTGGCAGC AATGTTATTT CCGATAATGG CACCCAGATT
                301
                       CAAGCAGGCA ATCATGTTCG CATTGGTACA ACCCAAACTC AAAGCCAAAG
                351
                       CGAAACCTAT CATCAAACCC AGAAATCAGG ATTGATGAGT GCAGGTATCG
                401
                       GCTTCACTAT TGGCAGCAAG ACAAACACAC AAGAAAACCA ATCCCAAAGC
25
                451
                       AACGAACATA CAGGCAGTAC CGTAGGCAGC TTGAAAGGCG ATACCACCAT
                501
                       TGTTGCAGGC AAACACTACG AACAAATCGG CAGTACCGTT TCCAGCCCGG
                551
                       AAGGCAACAA TACCATCTAT GCCCAAAGCA TAGACATTCA AGCGGCACAC
                601
                       AACAAATTAA ACAGTAATAC CACCCAAACC TATGAACAAA AAGG.CTAAC
                651
                       GGTGGCATTC AGTTCGCCCG TTACCGATTT GGCACAACAA ...
```

30 This corresponds to the amino acid sequence <SEQ ID 518; ORF117>:

```
1 .SGNNLNAKAA EVSSANGTLA VSANNDINIS AGINTTHVDD ASKHTGRSGG
51 GNKLVITDKA QSHHETAQSS TFEGKQVVLQ AGNDANILGS NVISDNGTQI
101 QAGNHVRIGT TQTQSQSETY HQTQKSGLMS AGIGFTIGSK TNTQENQSQS
151 NEHTGSTVGS LKGDTTIVAG KHYEQIGSTV SSPEGNNTIY AQSIDIQAAH
201 NKLNSNTTQT YEQKXLTVAF SSPVTDLAQQ ...
```

Computer analysis of this amino acid sequence gave the following results:

Homology with the pspA putative secreted protein of *N. meningitidis* (accession number AF030941) ORF117 and pspA protein show 45% aa identity in 224aa overlap:

```
Orf117: 4
                       NLNAKAAEVSSANGTLAVSANNDINISAGINTTHVDDASKHTGRSGGGNKLVITDKAOSH 63
40
                          +AAEV S G L ++A DI + AG
                                                      T +DA K+TGRSGGG K +T
                  1173 DIRIRAAEVGSEQGRLKLAAGRDIKVEAGKAHTETEDALKYTGRSGGGIKQKMTRHLKNQ 1232
          pspA:
          Orf117: 64
                       HETAQSSTFEGKQVVLQAGNDANILGSNVISDNGTQIQAGNHVRIGTTQTQSQSETYHOT 123
                       + A S T +GK+++L +G D + GSN+I+DN T + A N++ +
45
          pspA:
                  1233 NGQAVSGTLDGKEIILVSGRDITVTGSNIIADNHTILSAKNNIVLKAAETRSRSAEMNKK 1292
          Orf117: 124
                       QKSGLM-SAGIGFTIGSKTNTQENQSQSNEHTGSTVGSLKGDTTIVAGKHYEQIGSTVSS 182
                       +KSGLM S GIGFT GSK +TQ N+S++ HT S VGSL G+T I AGKHY Q GST+SS
                  1293 EKSGLMGSGGIGFTAGSKKDTQTNRSETVSHTESVVGSLNGNTLISAGKHYTQTGSTISS 1352
          pspA:
50
          Orf117: 183
                       PEGNNTIYAQSIDIQAAHNKLNSNTTQTYEOKXLTVAFSSPVTD 226
                       P+G+ I + I I AA N+ + + Q YEQK +TVA S PV +
                  1353 PQGDVGISSGKISIDAAQNRYSQESKQVYEQKGVTVAISVPVVN 1396
          pspA:
```

#### Homology with a predicted ORF from N.gonorrhoeae

ORF117 shows 90% identity over a 230aa overlap with a predicted ORF (ORF117ng) from N.gonorrhoeae:

5	orf117.pep	SGNNLNAKAAEVSSANGTLAVSANNDINIS	30
	orf117ng	IHFDADNHTIRGSTNEVGSSIQTKGDVTLLSGNNLNAKAAEVGSAKGTLAVYAKNDITIS	480
	orf117.pep	AGINTTHVDDASKHTGRSGGGNKLVITDKAQSHHETAQSSTFEGKQVVLQAGNDANILGS:  :::	90
10	orfl17ng	SGIHAGQVDDASKHTGRSGGGNKLVITDKAQSHHETAQSSTFEGKQVVLQAGNDANILGS	540
	orfl17.pep	NVISDNGTQIQAGNHVRIGTTQTQSQSETYHQTQKSGLMSAGIGFTIGSKTNTQENQSQS	150
15	orf117ng		600
	orf117.pep	NEHTGSTVGSLKGDTTIVAGKHYEQIGSTVSSPEGNNTIYAQSIDIQAAHNKLNSNTTQT	210
	orf117ng	NEHTGSTVGSLKGDTTIVASKHYEQTGSNVSSPEGNNLISTQSMDIGAAQNQLNSKTTQT	660
20	orf117.pep	YEQKXLTVAFSSPVTDLAQQ	230
	orfl17ng	YEQKGLTVAFSSPVTDLAQQAIAVAHKAAKQFDKAKTTALMPWRLPMQVGRLFKQAKAPK	720

An ORF117ng nucleotide sequence <SEQ ID 519> was predicted to encode a protein having amino acid sequence <SEQ ID 520>:

```
25
                    ..LLVQTEKDGL HNEQTFGEKK VFSENGKLHN YWRARRKGHD ETGHREONYT
                       LPEEITRDIS LGSFAYESHS KALSRHAPSQ GTELPQSNRD NIRTAKSNGI
                101
                       SLPYTPNSFT PLPGSSLYII NPANKGYLVE TDPRFANYRQ WLGSDYMLGS
                       LKLDPNNLHK RLGDGYYEQR LINEQIAELT GHRRLDGYQN DEEQFKALMD
                151
                201
                       NGATAARSMN LSVGIALSAE QAAQLTSDIV WLVQKEVKLP DGGTQTVLMP
30
                251
                       QVYVRVKNGG IDGKGALLSG SNTQINVSGS LKNSGTIAGR NALIINTDTL
                301
                       DNIGGRIHAQ KSAVTATQDI NNIGGILSAE QTLLLNAGNN INNQSTAKSS
                351
                       QNAQGSSTYL DRMAGIYITG KEKGVLAAQA GKDINIIAGQ ISNQSDQGQT
                       RLQAGRDINL DTVQTGKYQE IHFDADNHTI RGSTNEVGSS IQTKGDVTLL
                401
                451
                       SGNNLNAKAA EVGSAKGTLA VYAKNDITIS SGIHAGQVDD ASKHTGRSGG
35
                501
                       GNKLVITDKA QSHHETAQSS TFEGKQVVLQ AGNDANILGS NVISDNGTRI
                551
                       QAGNHVRIGT TQTQSQSETY HQTQKSGLMS AGIGFTIGSK TNTQENQSQS
                601
                       NEHTGSTVGS LKGDTTIVAS KHYEQTGSNV SSPEGNNLIS TQSMDIGAAQ
                651
                       NQLNSKTTQT YEQKGLTVAF SSPVTDLAQQ AIAVAHKAAK QFDKAKTTAL
                701
                       MPWRLPMQVG RLFKQAKAPK K*
```

#### 40 Further work revealed the following gonococcal partial DNA sequence <SEQ ID 521>:

	1	TTGCTTGTGC	AAACAGAAAA	AGACGGTTTG	CATAACGAGC	AAACCTTTGG
	51	CGAGAAGAAA	GTCTTCAGCG	AAAATGGTAA	GTTGCACAAC	TACTGGCGTG
	101	CGCGTCGTAA	AGGACATGAT	GAAACAGGGC	<b>ATCGTGAACA</b>	AAATTATACT
	151	TTGCCGGAGG	AAATCACACG	CGACATTTCA	CTGGGTTCAT	TTGCCTATGA
45	201	ATCGCATAGC	AAAGCATTAA	GCCGTCATGC	GCCCAGCCAA	GGCACTGAGT
	251	TGCCACAAAG	TAACCGGGAT	<b>AATATCCGTA</b>	CTGCGAAAAG	CAACGGTATT
	301	TCGCTACCCT	ATACGCCCAA	TTCTTTTACC	CCATTACCCG	GCAGCAGCTT
	351	ATACATTATC	AATCCTGCCA	ATAAAGGCTA	TCTTGTTGAA	ACCGATCCAC
	401	GCTTTGCCAA	CTACCGTCAA	TGGTTGGGTA	GTGACTATAT	GCTGGGCAGC
50	451	CTCAAACTAG	ACCCAAACAA	TTTACATAAA	CGTTTGGGTG	ATGGTTATTA
	501	CGAGCAACGT	TTAATCAATG	AACAAATCGC	AGAGCTGACA	GGGCATCGTC
	551	GTTTAGACGG	TTATCAAAAC	GACGAAGAAC	AATTTAAAGC	CTTAATGGAT
	601	AATGGCGCGA	CTGCGGCACG	TTCGATGAAT	CTCAGCGTTG	GCATTGCATT
	651	AAGTGCCGAG	CAAGCAGCGC	<b>AACTGACCAG</b>	CGATATTGTT	TGGTTGGTAC
55	701	AAAAAGAAGT	TAAACTTCCT	GATGGCGGCA	CACAAACCGT	ATTGATGCCA
	751	CAGGTTTATG	TACGCGTTAA	AAATGGCGGC	ATAGACGGTA	AAGGTGCATT
	801	GTTGTCAGGC	AGCAATACAC	AAATCAATGT	TTCAGGCAGC	CTGAAAAACT
	851	CAGGCACGAT	TGCAGGGCGC	AATGCGCTTA	TTATCAATAC	CGATACGCTA
	901	GACAATATCG	GTGGGCGTAT	TCATGCGCAA	AAATCAGCGG	TTACGGCCAC
60	951	ACAAGACATC	AATAATATTG	GCGGCATTCT	TTCTGCCGAA	CAGACATTAT
	1001	TGCTCAATGC	GGGTAACAAC	ATCAACAACC	AAAGCACGGC	CAAGAGCAGT
	1051	CAAAATGCAC	AAGGTAGCAG	CACCTACCTA	GACCGAATGG	CAGGTATTTA

40

```
1101
                     TATCACAGGC AAAGAAAAG GTGTTTTAGC AGCGCAGGCA GGCAAAGACA
                     TCAACATCAT TGCCGGTCAA ATCAGCAATC AATCAGATCA AGGGCAAACC
               1151
               1201
                     CGGCTGCAGG CAGGACGCGA CATTAACCTG GATACGGTAC AAACCGGCAA
               1251
                     ATATCAAGAA ATCCATTTTG ATGCCGATAA CCATACCATC CGAGGTTCAA
5
                     CGAACGAAGT CGGCAGCAGC ATTCAAACAA AAGGCGATGT TACCCtatTG
               1301
                     TCAGGGAATA ATCTCAATGC CAAAGCTGCC GAAGTCGGCA GCGCAAAAGG
               1351
               1401
                     CACACTTGCC GTGTATGCTA AAAATGACAT TACTATCAGC TCAGGCATCC
                     ATGCCGGCCA AGTTGATGAT GCGTCCAAAC ATACAGGCAG AAGCGGCGGC
               1451
               1501
                     GGTAATAAAT TAGTCATTAC CGATAAAGCC CAAAGTCATC ACGAAACTGC
                     TCAAAGCAGC ACCTTTGAAG GCAAGCAAGT TGTATTGCAG GCAGGAAACG
10
               1551
               1601
                     ATGCCAACAT CCTTGGCAGT AATGTTATTT CCGATAATGG CACCCGGATT
                     CAAGCAGGCA ATCATGTTCG CATTGGTACA ACCCAAACTC AAAGCCAAAG
               1651
               1701
                     CGAAACCTAT CATCAAACCC AAAAATCAGG ATTGATGAGT GCAGGTATCG
                     GCTTCACTAT TGGCAGCAAG ACAAACACAC AAGAAAACCA ATCCCAAAGC
               1751
15
                     AACGAACATA CAGGCAGTAC CGTAGGCAGC CTGAAAGGCG ATACCACCAT
               1801
               1851
                     TGTTGCAAGC AAACACTACG AACAAACCGG CAGCAACGTT TCCAGCCCTG
                     AGGGCAACAA CCTTATCAGC ACGCAAAGTA TGGATATTGG CGCAGCACAA
               1901
                     AACCAATTAA ACAGCAAAAC CACCCAAACC TACGAACAAA AAGGCTTAAC
               1951
                     GGTGGCATTC AGTTCGCCCG TTACCGATTT GGCACAACAA GCGATTGCCG
TAGCACACAA AGCAGCAAAC AAGTCGGACA AAGCAAAAAC GACCGCGTTA
               2001
20
               2051
                     ATGCCATGGC GGCTGCCAAT GCAGGTTGGC AGGCCTATCA AACAGGCAAA
               2101
               2151 GGCGCACAAA ACTTAG
```

This corresponds to the amino acid sequence <SEQ ID 522; ORF117ng-1>:

```
LLVQTEKDGL HNEQTFGEKK VFSENGKLHN YWRARRKGHD ETGHREQNYT
25
                 51
                      LPEEITRDIS LGSFAYESHS KALSRHAPSQ GTELPQSNRD NIRTAKSNGI
                      SLPYTPNSFT PLPGSSLYII NPANKGYLVE TDPRFANYRQ WLGSDYMLGS
                101
                     LKLDPNNLHK RLGDGYYEQR LINEQIAELT GHRRLDGYQN DEEQFKALMD
                151
                201 NGATAARSMN LSVGIALSAE QAAQLTSDIV WLVQKEVKLP DGGTQTVLMP
                251
                     QVYVRVKNGG IDGKGALLSG SNTQINVSGS LKNSGTIAGR NALIINTDTL
30
                      DNIGGRIHAQ KSAVTATQDI NNIGGILSAE QTLLLNAGNN INNQSTAKSS
                301
                      QNAQGSSTYL DRMAGIYITG KEKGVLAAQA GKDINIIAGQ ISNQSDQGQT
                351
                      RLQAGRDINL DTVQTGKYQE IHFDADNHTI RGSTNEVGSS IQTKGDVTLL
                401
                451
                      SGNNLNAKAA EVGSAKGTLA VYAKNDITIS SGIHAGQVDD ASKHTGRSGG
                      GNKLVITDKA QSHHETAQSS TFEGKQVVLQ AGNDANILGS NVISDNGTRI
QAGNHVRIGT TQTQSQSETY HQTQKSGLMS AGIGFTIGSK TNTQENQSQS
                501
35 .
                 551
                 601
                      NEHTGSTVGS LKGDTTIVAS KHYEQTGSNV SSPEGNNLIS TQSMDIGAAQ
                 651
                      NQLNSKTTQT YEQKGLTVAF SSPVTDLAQQ AIAVAHKAAN KSDKAKTTAL
                      MPWRLPMQVG RPIKQAKAHK T*
```

ORF117ng-1 shows the same 90% identity over a 230aa overlap with ORF117. In addition, it shows homology with a secreted *N.meningitidis* protein in the database:

```
gi|2623258 (AF030941) putative secreted protein [Neisseria meningitidis]Length = 2273
```

Score = 604 bits (1541), Expect = e-172 Identities = 325/678 (47%), Positives = 449/678 (65%), Gaps = 22/678 (3%) 45 LLVQTEKDGLHNEQTFGEKKVFSENGKLHNYWRARRKGHDETGHREQNYTLPEEITRDIS 60 L+V T + L N++T G K + ++ G LH Y R +KG D TG+ E++ T LIVGTPESALDNDETLGTKTI-TDKGDLHRYHRHHKKGRDSTGYSRSPYEPAPEVS-SIR 796 Sbjct: 739 50 LGSFAYESHSKALSRHAPSQGTELPQSNRDNIRTAKSNGISLPYTPNSFTPLPGSSLYII 120 Query: 61 +G AY+ + AP Q +++P + NGI MGISAYKGY------APQQASDIPGTV---VPVVAENGIHPTFT-----LPNSSLFAI 840 Sbjct: 797 NPANKGYLVETDPRFANYRQWLGSDYMLGSLKLDPNNLHKRLGDGYYEQRLINEQIAELT 180 Query: 121 55 P NKGYL+ETDP F +YR+WLGS YML +L+ DPN++HKRLGDGYYEQ+L+NEQIA+LT APNNKGYLIETDPAFTDYRKWLGSGYMLAALQQDPNHIHKRLGDGYYEQKLVNEQIAKLT 900 Sbjct: 841 GHRRLDGYQNDEEQFKALMDNGATAARSMNLSVGIALSAEQAAQLTSDIVWLVQKEVKLP 240 Query: 181 G+RRLDGY NDEEQFKALMDNG T A+ + L+ GIALSAEQ A+LTSDIVWL + V LP 60 GYRRLDGYTNDEEQFKALMDNGITIAKELQLTPGIALSAEQVARLTSDIVWLENETVTLP 960 Query: 241 DGGTQTVLMPQVYVRVKNGGIDGKGALLSGSNTQINVSGSLKN-SGTIAGRNALIINTDT 299 DG TOTVL P+VYVR + ++G+GALLSGS I SG+++N G IAGR ALI+N Sbjct: 961 DGTTQTVLKPKVYVRARPKDMNGQGALLSGSVVDIG-SGAIENRGGLIAGREALILNAQN 1019 65 Query: 300 LDNIGGRIHAQKSAVTATQDINNIGGILSAEQTLLLNAGNNINNQSTAKSSQNAQGSSTY 359

```
A DINGIAE LLL ANNI ++S +S+QN QGS
                      + N+ G + +
          Sbjct: 1020 IKNLQGDLQGKNIFAAAGSDITNTGSI-GAENALLLKASNNIESRSETRSNQNEQGSVRN 1078
          Query: 360 LDRMAGIYITGKEKGVLAAQAGKDINIIAGQISNQSDQGQTRLQAGRDINLDTVQTGKYQ 419
5
                      + R+AGIY+TG++ G +
                                         AG +I + A +++NQS+ GQT L AG DI DT
          Sbjct: 1079 IGRVAGIYLTGRQNGSVLLDAGNNIVLTASELTNQSEDGQTVLNAGGDIRSDTTGISRNQ 1138
          Query: 420 EIHFDADNHTIRGSTNEVGSSIQTKGDVTLLSGNNLNAKAAEVGSAKGTLAVYAKNDITI 479
                                   NEVGS+I+T+G+++L + ++ +AAEVGS +G L + A DI +
                         FD+DN+ IR
          Sbjct: 1139 NTIFDSDNYVIRKEQNEVGSTIRTRGNLSLNAKGDIRIRAAEVGSEQGRLKLAAGRDIKV 1198
10
          Query: 480 SSGIHAGQVDDASKHTGRSGGGNKLVITDKAQSHHETAQSSTFEGKQVVLQAGNDANILG 539
                             + +DA K+TGRSGGG K +T
                                                    ++ + A S T +GK+++L +G D + G
          Sbjct: 1199 EAGKAHTETEDALKYTGRSGGGIKQKMTRHLKNQNGQAVSGTLDGKEIILVSGRDITVTG 1258
15
                      SNVISDNGTRIQAGNHVRIGTTQTQSQSETYHQTQKSGLM-SAGIGFTIGSKTNTQENQS 598
          Query: 540
                      SN+I+DN T + A N++ + +T+S+S
                                                     ++ +KSGLM S GIGFT GSK +TQ N+S
           Sbjct: 1259 SNIIADNHTILSAKNNIVLKAAETRSRSAEMNKKEKSGLMGSGGIGFTAGSKKDTQTNRS 1318
                      QSNEHTGSTVGSLKGDTTIVASKHYEQTGSNVSSPEGNNLISTQSMDIGAAQNQLNSKTT 658
20
                       ++ HT S VGSL G+T I A KHY QTGS +SSP+G+ IS+ + I AAQN+ + ++
           Sbjct: 1319 ETVSHTESVVGSLNGNTLISAGKHYTQTGSTISSPQGDVGISSGKISIDAAQNRYSQESK 1378
           Query: 659 QTYEQKGLTVAFSSPVTD 676
25
                       Q YEQKG+TVA S PV +
           Sbjct: 1379 QVYEQKGVTVAISVPVVN 1396
```

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 63

30 The following partial DNA sequence was identified in N.meningitidis <SEQ ID 523>:

```
1 ATGATTTACA TCGTACTGTT TCTAGCTGTC GTCCTCGCCG TTGTCGCCTA
51 CAACATGTAT CAGGAAAACC AATACCGCAA AAAAGTGCGC GACCAGTTCG
101 GACACTCCGA CAAAGATGCC CTGCTCAACA GCAWAACCAG CCCTGCTCGCC
151 GACGGCAAAC CGTCCGGCGG GTCAGTCATG ATGCCGAAAC CCCAACCGGC
351 AACAGGATGC CGTCTACATC GCCAAGCAGA AACAGGCAAA AGCCTCCCCG
301 TTCAAAACCG AAATCGAAAC CGCCTTGGAA GAAAGCGGCA TTATCGGCAA
351 CTCCGCCCAC ACCGTTTCCG AACCCCAAAC CGGACATTCC GCAACGAAAC
401 CTGCCGACGC GTCGGCAAAA CCTGCACCG TTCCGCAAAC ACCTGCAAAA
451 CCGCTGATTA CGCTCAAAGA ACTGTCAAAA GTCGAAATTAT CCTGGTTTGA
501 CGTGCGCATC GACTTCATCT CCTAT...
```

This corresponds to the amino acid sequence <SEQ ID 524; ORF119>:

```
1 MIYIVLFLAV VLAVVAYNMY QENQYRKKVR DQFGHSDKDA LLNSXTSHVR
51 DGKPSGGSVM MPKPQPAVKK TAKPQDPXMR NLQEQDAVYI AKQKQAKASP
45 101 FKTEIETALE ESGIIGNSAH TVSEPQTGHS ATKPADASAK PAPVPQTPAK
151 PLITLKELSK VELSWFDVRI DFISY...
```

Further work revealed the complete nucleotide sequence <SEQ ID 525>:

	1	ATGATTTACA	TCGTACTGTT	TCTAGCTGTC	GTCCTCGCCG	TTGTCGCCTA
	51	CAACATGTAT	CAGGAAAACC	AATACCGCAA	AAAAGTGCGC	GACCAGTTCG
50	101	GACACTCCGA	CAAAGATGCC	CTGCTCAACA	GCAAAACCAG	CCATGTCCGC
	151	GACGGCAAAC	CGTCCGGCGG	GTCAGTCATG	ATGCCGAAAC	CCCAACCGGC
	201	GGTCAAAAAA	ACGGCAAAAC	CCCAAGACCC	CGCCATGCGC	AACCTGCAAG
	251	AACAGGATGC	CGTCTACATC	GCCAAGCAGA	AACAGGCAAA	AGCCTCCCCG
	301	TTCAAAACCG	AAATCGAAAC	CGCCTTGGAA	GAAAGCGGCA	TTATCGGCAA
55	351	CTCCGCCCAC	ACCGTTTCCG	AACCCCAAAC	CGGACATTCC	GCACCGAAAC
	401	CTGCCGACGC	GCCGGCAAAA	CCTGCACCCG	TTCCGCAAAC	ACCTGCAAAA
	451	CCGCTGATTA	CGCTCAAAGA	ACTGTCAAAA	GTCGAATTAC	CCTGGTTTGA
	501	CGTGCGCTTC	GACTTCATCT	CCTATATCGC	GCTGACCGAA	GCCAAAGAAC
	551	TGCACGCACT	GCCGCGCCTT	TCCAACCGCT	GCCGCTACCA	GATTGTCGGC
60	601	TGCACCATGG	ACGACCATTT	CCAGATTGCC	GAACCCATCC	CGGGCATCCG

	•		•		•					
	651	CTATCAGGCA	TTTATCGTGG	GTATTCAGGC	AGTCAGCCGC	AACGGACTTG				
	701	CCTCGCAGGA	AGAACTCTCC	GCATTCAACC	GCCAGGTGGA	CGCATTCGCA				
	751	CAAAGCATGG	GCGGTCAGAC	GCTGCACACC	GACCTTGCCG	CCTTTATCGA				
	· 801	AGTGGCTTCC	GCACTGGACG	CATTCTGCGC	GCGCGTCGAC	CAGACCATCG				
5	851	CCATCCATTT	GGTTTCCCCG	ACCAGCATCA	GCGGCGTAGA	ACTGCGTTCC				
	901	GCCGTAACGG	GCGTGGGTTT	<b>CGTTTTGGAA</b>	GACGACGGCG	CGTTCCACTA				
	951	TACCGACACG	TCGGGCTCGA	CCATGTTCTC	CATCTGCTCG	CTCAACAACG				
	1001	AGCCGTTTAC	CAACGCCCTT	TTGGACAACC	AGTCCTACAA	AGGCTTCAGT				
	1051	ATGCTGCTCG	ACATCCCGCA	CTCTCCGGCA	GGCGAAAAAA	CCTTCGACGA				
10	1101	TTTGTTTATG	GATTTGGCGG	TACGCCTGTC	CGGCCAGTTG	AACCTGAATC				
	1151	TGGTCAACGA	CAAAATGGAA	GAAGTTTCGA	CCCAATGGCT	CAAAGACGTG				
	1201	CGCACTTATG	TATTGGCGCG	TCAGTCCGAG	ATGCTCAAAG	TCGGTATCGA				
	1251	ACCGGGCGGC	AAAACCGCAT	TGCGCCTGTT	CTCCTAA					
This corresponds to the amino acid sequence <seq 526;="" id="" orf119-1="">:</seq>										
15	1	MIYIVLFLAV	VLAVVAYNMY	QENQYRKKVR	DOFGHSDKDA	LLNSKTSHVR				
	C 1			mawnoona.						

15	1	MIYIVLFLAV	VLAVVAYNMY	QENQYRKKVR	DQFGHSDKDA	LLNSKTSHVR
	51	DGKPSGGSVM	MPKPQPAVKK	TAKPQDPAMR	NLQEQDAVYI	AKQKQAKASP
	101				APKPADAPAK	
	151	PLITLKELSK	VELPWFDVRF	DFISYIALTE	AKELHALPRL	SNRCRYQIVG
	201				NGLASQEELS	
20	251	QSMGGQTLHT	DLAAFIEVAS	ALDAFCARVD	QTIAIHLVSP	<b>TSISGVELRS</b>
	301				LNNEPFTNAL	
	351	MLLDIPHSPA	GEKTFDDLFM	DLAVRLSGQL	NLNLVNDKME	EVSTQWLKDV
	401	RTYVLARQSE	MLKVGIEPGG	KTALRLFS*		

Computer analysis of this amino acid sequence gave the following results:

## 25 Homology with a predicted ORF from N.meningitidis (strain A)

ORF119 shows 93.7% identity over a 175aa overlap with an ORF (ORF119a) from strain A of N. meningitidis:

30	orf119.pep	10 MIYIVLFLAVVLA	20 VVAYNMYQENÇ	30 YRKKVRDQF(	40 GHSDKDALLNS	50 SXTSHVRDGK	60 PSGGSVM
	orf119a	MIYIVLFLAAVLA	VVAYNMYQENC	YRKKVRDQF	GHSDKDALLNS	SKTSHVRDGK	PSGGPVM
		10	20	30	40	50	60
0.5		70	80	90	100	110	120
35	orf119.pep	MPKPQPAVKKTAK	PQDPXMRNLQE	QDAVYIAKQI	KQAKASPFKTE	CIETALEESG:	IIGNSAH
			111 111111	111111111	[		111111
	orf119a	MPKPQPAVKKTAK				CIETALEESG:	IIGNSAH
		70	80	90	100	110	120
40		. 120	140		4.00		
40		130	140	150	160	170	
	orf119.pep	TVSEPQTGHSATK	PADASAKPAPV	POTPAKPLI:	<b>PLKELSKVELS</b>	SWFDVRIDFI:	SY
			1111 111:13	311111111	111111111	111111111	H -
	orf119a	TVPEPQTGHSAPK	PADAPAKPVPV	POTPAKPLI	TLKELSKVELI	WFDVRFDFI:	SYIALTE
		130	140	150	160	170	180
45						• •	
	orfl19a	AKELHALPRLSNR	CRYQIVGCTME	DHFQIAEPI	PGIRYOAFIVO	SIOAVSRNGL	ASOEELS
		190	200	210	220	230	240

The complete length ORF119a nucleotide sequence <SEQ ID 527> is:

	1	ATGATTTACA	TCGTACTGTT	CCTCGCCGCC	GTCCTCGCCG	TTGTCGCCTA
50	51	CAATATGTAT	CAGGAAAACC	AATACCGCAA	AAAAGTGCGC	GACCAGTTCG
	101	GGCACTCCGA	CAAAGATGCC	CTGCTCAACA	GCAAAACCAG	CCATGTCCGC
	151	GACGGCAAAC	CGTCCGGCGG	GCCAGTCATG	ATGCCGAAAC	CCCAACCGGC
	201	GGTCAAAAAA	ACGGCAAAAT	CCCAAGACCC	CGCCATGCGC	AACCTGCAAG
:	251	AGCAGGATGC	CGTCTACATC	GCCAAGCAGA	AACAGGCAAA	AGCCTCCCCG
55	301	TTCAAAACCG	AAATCGAAAC	CGCCTTGGAA	GAAAGCGGCA	TTATCGGCAA
	351	CTCCGCCCAC	ACCGTTCCCG	AACCCCAAAC	CGGACATTCC	GCACCAAAAC
	401	CTGCCGACGC	GCCGGCAAAA	CCTGTTCCCG	TTCCGCAAAC	GCCGGCAAAA
	451				GTCGAGCTGC	
	501	CGTGCGCTTC	GACTTCATCT	CTTATATCGC	GCTGACCGAA	GCCAAAGAAC
60	551	TGCACGCACT	GCCGCGCCTT	TCCAACCGCT	GCCGCTACCA	GATTGTCGGC
	601	TGCACCATGG	ACGACCATTT	CCAGATTGCC	GAACCCATCC	CGGGCATCCG

			TTTATCGTGG					
			AGAACTCTCC					
			GCGGTCAGAC					
_			GCACTGGACG					
5			GGTTTCCCCG					
			GCGTGGGTTT					
			TCGGGCTCGA					
			CAATGCCCTT					
10			ACATCCCGCA					
10			GATTTGGCGG CAAAATGGAA					
			TATTGGCTCG					
			AAAACCGCAT			3 ICGGIAIC	.GA	
						0 <b>~.</b>		
	This encodes a prot	em naving	g amino acid	sequence <	2EQ ID 326	<b>5</b> /:		
15	·1 MI	YIVLFLAA	VLAVVAYNMY	QENQYRKKVR	DOFGHSDKD	A LLNSKTSH	<b>IV</b> R	
			MPKPQPAVKK					
			ESGIIGNSAH					
			VELPWFDVRF					
			EPIPGIRYQA					
20			DLAAFIEVAS					
			DDGAFHYTDT					
			GEKTFDDLFM		NENTANDKW	E EVSTQWLK	DV	
	401 RT	IVLARQSE	MLKVGIEPGG	KTALKLES*				
	ORF119a and ORF	119-1 sho	w 98.6% id	entity in 428	aa overlap:	:		
25			10	20	30	40	50	60
25	orf119a.pep	MIYIVI	LFLAAVLAVVA					
	222224P-F							
	orf119-1		FLAVVLAVVA					
			10	20	30	40	50	60
30								
			70	80	90	100	110	120
	orf119a.pep		PAVKKTAKSQD:					
25	orf119-1	MPKPQI	PAVKKTAKPQD					
35			70	80	90	100	110	120
	•		130	140	150	160	170	100
	orf119a.pep	TUDED	TGHSAPKPAD				170 WDENETGYT	180
	OIIIIJa.pep							
40	orf119-1		TGHSAPKPAD					
•••			130	140		160	170	180
			190	200	210	220	230	240
	orf119a.pep		ALPRLSNRCRY					
45								
	orī119-1	AKELH	ALPRLSNRCRY				_	
			190	200	210	220	230	240
			250	260	070	000	000	200
50	orfl19a.pep	N END O	250 VDAFAHSMGGQ			280	290	300
30	Offiliag.pep		:					
	orf119-1		VDAFAQSMGGQ					
	011115 1	HIMNY	250	260		280	290	300
			200				250	500
55			310	320	330	340	350	360
	orf119a.pep	AVTGV	GFVLEDDGAFH					
			11111111111					
	orf119-1	AVTGV	GFVLEDDGAFH	YTDTSGSTMFS	SICSLNNEPFT	NALLDNOSY	KGFSMLLDIP	HSPA
			310	320		340	350	360
60								
			370	380		400	410	420
	orf119a.pep	GEKTF	DDLFMDLAVRL	SGQLNLNLVNI	OKMEEVSTQWL	KDVRTYVLA	ROSEMLKVGI	EPGG
		11111						
65	orf119-1	GEKTF	DDLFMDLAVRL					
03			370	380	390	400	410	420

#### 5 Homology with a predicted ORF from N.gonorrhoeae

ORF119 shows 93.1% identity over a 175aa overlap with a predicted ORF (ORF119ng) from

### N.gonorrhoeae:

	orf119.pep	MIYIVLFLAVVLAVVAYNMYQENQYRKKVRDQFGHSDKDALLNSXTSHVRDGKPSGGSVM	60
10	orf119ng	MIYIVLFLAAVLAVVAYNMYQENQYRKKVRDQFGHSDKDALLNSKTSHVRDGKPSGGPVM	60
	orf119.pep	MPKPQPAVKKTAKPQDPXMRNLQEQDAVYIAKQKQAKASPFKTEIETALEESGIIGNSAH	120
15	orf119ng	MPKPQPAVKKPAKPQDSAMRNLQEQDAVYIAKQKQAKASPFKTEIETALEEIGIIGNSAH	120
	orf119.pep	TVSEPQTGHSATKPADASAKPAPVPQTPAKPLITLKELSKVELSWFDVRIDFISY	175
	orf119ng	TVSEPQTGHSAPKPADAPAKPVPVPQTPAKPLITLKELSKVELPWFDVRFDFISYIALTE	180

## The complete length ORF119ng nucleotide sequence <SEQ ID 529> is:

20	1	ATGATTTACA				
	• 51	CAATATGTAT	CAGGAAAACC	<b>AATACCGCAA</b>	AAAAGTGCGC	GACCAGTTCG
	101	GACACTCCGA	CAAAGATGCC	CTGCTCAACA	GCAAAACCAG	CCATGTCCGC
	151	GACGGCAAAC	CGTCCGGCGG	GCCAGTCATG	ATGCCGAAAC	CCCAACCGGC
	201	GGTCAAAAAA	CCGGCCAAAC	CCCAAGACTC	CGCCATGCGC	<b>AACCTGCAAG</b>
25	251	AACAGGATGC	CGTCTACATC	GCCAAGCAGA	AACAGGCAAA	AGCCTCCCCG
	301	TTCAAAACCG	AAATCGAAAC	CGCCTTGGAA	GAAATCGGCA	TTATCGGCAA
	351	CTCCGCCCAC	ACCGTTTCCG	AACCCCAAAC	CGGACATTCC	GCACCGAAAC
	401	CTGCCGACGC	GCCGGCAAAA	CCCGTTCCCG	TTCCGCAAAC	GCCGGCAAAA
	451	CCGCTGATTA	CGCTCAAAGA	GCTGTCGAAG	GTCGAGCTGC	CCTGGTTTGA
30	501	CGTGCGCTtc	<b>GACTTCATCT</b>	CCTATATCGC	GCTGACCGAA	GCCAAAGAAC
	551	TGCACGCACT	GCCGCGCCTT	tccAACCGCT	GCCGCTACCA	GATTGTCGGC
	601				GAACCCATCC	
	651				AGTCAGCCGC	
	701				GCCAGGCGGA	
35	751	CAAAGCATGG	GCGGTCAGAC	GCTGCACACC	GACCTTGCCG	CCTTTATCGA
	801	AGTGGCTTCC	GCACTGGACG	CATTCTGCGC	GCGCGTCGAC	CAGACCATCG
	851	CCATCCATTT	GGTTTCGCCG	ACCAGCATCA	GCGGCGTAGA	ACTGCGTTCC
	901	GCCGTAACGG	GCGTGGGTTT	CGTTTTGGAA	GACGACGGCG	CGTTCCACTA
40	951	TACCGACACG	TCGGGCTCGA	CCATGTTCTC	CATCTGCTCG	CTCAACAACG
40	1001	AGCCGTTTAC	CAATGCCCTT	TTGGACAACC	AGTCCTACAA	AGGCTTCAGT
	1051				GGCGAAAAA	
	1101				CGGTCAGTTG	
	1151	TGGTCAACGA	CAAAATGGAA	GAAGTTTCGA	CCCAATGGCT	CAAAGACGTA
	1201	CGCACTTATG	TATTGGCGCG	TCAGTCCGAG	ATGCTCAAAG	TCGGTATCGA
45	1251	ACCGGGCGGC	AAAACCGCCC	TGCGCCTGTT	TTCATAA	

## This encodes a protein having amino acid sequence <SEQ ID 530>:

	. 1	MIYIVLFLAA	VLAVVAYNMY	<b>QENQYRKKVR</b>	DOFGHSDKDA	LLNSKTSHVR
	51	DGKPSGGPVM	MPKPQPAVKK	PAKPQDSAMR	NLQEQDAVYI	AKQKQAKASP
	101	FKTEIETALE	EIGIIGNSAH	TVSEPQTGHS	APKPADAPAK	PVPVPQTPAK
50	151	PLITLKELSK	VELPWFDVRF	DFISYIALTE	AKELHALPRL	SNRCRYQIVG
	201	CTMDDHFQIA	EPIPGIRYQA	FIVGIQAVSR	NGLASQEELS	AFNRQADAFA
	251	QSMGGQTLHT	DLAAFIEVAS	ALDAFCARVD	QTIAIHLVSP	TSISGVELRS
	301				LNNEPFTNAL	
	351	MLLDIPHSPA	GEKTFDDLFM	DLAVRLSGQL	NLNLVNDKME	EVSTQWLKDV
55	401	RTYVLARQSE	MLKVGIEPGG	KTALRLFS*		_

### ORF119ng and ORF119-1 show 98.4% identity over 428 aa overlap:

		10	20	30	40	50	60
	orfl19ng	MIYIVLFLAAVLAVV	AYNMYQEN	QYRKKVRDQFGI	HSDKDALLNS	KTSHVRDGKP	SGGPVM
<b>60</b>		111111111111					
60	orf119-1	MIYIVLFLAVVLAVV	'Aynmyqen	QYRKKVRDQFGI	HSDKDALLNS	KTSHVRDGKP	SGGSVM
		10	20	30	40	50	60

-311-

		70	80	90	100	110	120
	orf119ng	MPKPQPAVKKPAKP(	ODSAMRNLQI	EQDAVYIAKQK	<b>QAKASPFKTE</b>	IETALEEIGI:	IGNSAH 
		[]]][]]					TGNSAH
5	orf119-1	MPKPQPAVKKTAKP(	ODPAMRNLQI 80	EODAVATAKOK 90	QAKASPIKIE 100	110	120
		70	80	90	100		
		130	140	150	160	170	180
	orfl19ng	THE POTCHEN DED	ΑΠΑΡΑΚΡΟΡ	VPQTPAKPLIT	LKELSKVELE	WFDVRFDFIS	YIALTE
10	011113119		111111111	11:11:11:11		1111111111	11111
10	orf119-1	TVSEPQTGHSAPKP	ADAPAKPAP	VPQTPAKPLI1	LKELSKVELI 160	170	180
		130	140	150	160	170	100
		190	200	210	220	230	240
1.5	orf119ng	APETUAL DELCARC	RYOTVGCTM	DDHFOIAEPII	GIRYQAFIV	SIQAVSRNGLA	SQEELS
15	oriling		111111111	11111111111			111111
	orf119-1	AKELHALPRLSNRC	RYQIVGCTM	IDDHFQIAEPII	GIRYQAFIV	<b>SIQAVSRNGLA</b>	COEETS
	011111	190	200	210	220	230	240
			0.60	270	280	290	300
20		250 AFNRQADAFAQSMG	260	2/U (AETEVASAT.D)			
	orf119ng	AFNRQADAFAQSMG	TILLIIII MULTILIIII	METEANSWID			111111
	orf119-1	AFNRQVDAFAQSMG	GOTLHTDLA	AFIEVASALD	AFCARVDQTI	AIHLVSPTSIS	SGAETES
	OLITIA-I	250	260	270	280	290	300
25						250	360
23		310	320	330	340	350	
	orf119ng	AVTGVGFVLEDDGA	AFHYTDTSG	STMESICSENN	FELLUATION	021KGF3MDD	IIIIIII
	ca.1.01	AVTGVGFVLEDDG		STMFSICSLNN	EPFTNALLDN	QSYKGFSMLL!	DIPHSPA
30	orf119-1	310	320	330	340	350	360
30							400
		370	380	390	400	410	420
	orf119ng	GEKTFDDLFMDLA	VRLSGQLNL	NLVNDKMEEVS	TOMPKDAKLI	VLARQSEMLA	1111111
		GEKTFDDLFMDLA	HIIIIIII		TOWT.KDVRTY	VLAROSEMLK	VGIEPGG
35	orf119-1	GERTEDDLEMDIA	380	390	400	410	420
		3.0	-				
		429					
	orf119ng	KTALRLFSX					
40							
	orf119-1	KTALRLFSX					
				~	1		ha aanaa

Based on this analysis, including the presence of a putative leader sequence in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

## 45 Example 64

The following partial DNA sequence was identified in N.meningitidis <SEQ ID 531>

50	1 51 101 151 201 251 301 351	CATCTCCCTC	GAAAGCACCA TTCATTGGTA TTACCGAGCG CGCGGCAATA CATCGGCGGT TCAATCATTT ATCGGCGCGG	CCGGTACGAT GTCGGCGGCA CACCAAAGAA TTTYGCAGCA TTGGTCGGCG TGTAACCGAC TCGCCTGTTC	GAAGCTGCTG TCGGCGTGAT ATCGGCATAC GTTTTTGATT TGGGTTTGTC TTCCCGATGG GACCGGAATC	ATTTCCTCA GAACATCATG GGATGGCAAT GAGGCGGTGT CGCCGCCGTC ACATTTCCGC GGCATCGCGT
55	401 451	TCGGCTTTAT TTGGCACAGG		AAAGCAGCCA	AACTCAATCC	GATAGACGCA

This corresponds to the amino acid sequence <SEQ ID 532; ORF134>:

51 LVSVTERTKE IGI	SDXIRQIV ESTTGTMKLL ISSIALISLV VGGIGVMNIM IRMAIGAR RGNIXQQFLI EAVLICVIGG LVGVGLSAAV MDISAMSV IGAVACSTGI GIAFGFMPAN KAAKLNPIDA
-------------------	---

```
Further work revealed the complete nucleotide sequence <SEQ ID 533>:
```

```
ATGTCGGTGC AAGCAGTATT GGCGCACAAA ATGCGTTCGC TTCTGACGAT
                     GCTCGGCATC ATCATCGGTA TCGCGTCGGT GGTTTCCGTC GTCGCATTGG
                 51
                101
                     GCAATGGTTC GCAGAAAAA ATCCTTGAAG ACATCAGTTC GATAGGGACG
5
                     AACACCATCA GCATCTTCCC GGGGCGCGGC TTCGGCGACA GGCGCAGCGG
                151
                201
                     CAGGATTAAA ACCCTGACCA TAGACGACGC AAAAATCATC GCCAAACAAA
                     GCTACGTTGC TTCCGCCACG CCCATGACTT CGAGCGGCGG CACGCTGACT
                251
                301
                     TACCGCAACA CCGACCTGAC CGCCTCGCTT TACGGCGTGG GCGAACAATA
                     TTTCGACGTG CGCGGACTGA AGCTGGAAAC GGGGCGGCTG TTTGACGAAA
                351
10
                401
                     ACGATGTGAA AGAAGACGCG CAGGTCGTCG TCATCGACCA AAATGTCAAA
                     GACAAACTCT TTGCGGACTC GGATCCGTTG GGTAAAACCA TTTTGTTCAG
                451
                501
                     GAAACGCCCC TTGACCGTCA TCGGCGTGAT GAAAAAGAC GAAAACGCTT
                     TCGGCAATTC CGACGTGCTG ATGCTTTGGT CGCCCTATAC GACGGTGATG
                551
                601
                     CACCAAATCA CAGGCGAGAG CCACACCAAC TCCATCACCG TCAAAATCAA
15
                651 AGACAATGCC AATACCCAGG TTGCCGAAAA AGGGCTGACC GATCTGCTCA
                701 AAGCGCGGCA CGGCACGGAA GATTTCTTCA TGAACAACAG CGACAGCATC
                751
                     AGGCAGATAG TCGAAAGCAC CACCGGTACG ATGAAGCTGC TGATTTCCTC
                     CATCGCCCTG ATTTCATTGG TAGTCGGCGG CATCGGCGTG ATGAACATCA
                801
                     TGCTGGTGTC CGTTACCGAG CGCACCAAAG AAATCGGCAT ACGGATGGCA
                851
20
                     ATCGGCGCGC GGCGCGCAA TATTTTGCAG CAGTTTTTGA TTGAGGCGGT
                901
                951
                     GTTAATCTGC GTCATCGGCG GTTTGGTCGG CGTGGGTTTG TCCGCCGCCG
                     TCAGCCTCGT GTTCAATCAT TTTGTAACCG ACTTCCCGAT GGACATTTCC GCCATGTCCG TCATCGGCGC GGTCGCCTGT TCGACCGGAA TCGGCATCGC
               1001
               1051
                     GTTCGGCTTT ATGCCTGCCA ATAAAGCAGC CAAACTCAAT CCGATAGACG
               1101
25
               1151
                     CATTGGCACA GGATTGA
```

This corresponds to the amino acid sequence <SEQ ID 534; ORF134-1>:

```
1 MSVQAVLAHK MRSLLTMLGI IIGIASVVSV VALGNGSQKK ILEDISSIGT
51 NTISIFPGRG FGDRRSGRIK TLTIDDAKII AKQSYVASAT PMTSSGGTLT
101 YRNTDLTASL YGVGEQYFDV RGLKLETGRL FDENDVKEDA QVVVIDQNVK
151 DKLFADSDPL GKTILFRKRP LTVIGVMKKD ENAFGNSDVL MLWSPYTTVM
201 HQITGESHTN SITVKIKDNA NTQVAEKGLT DLLKARHGTE DFFMNNSDSI
251 RQIVESTTGT MKLLISSIAL ISLVVGGIGV MNIMLVSVTE RTKEIGIRMA
301 IGARRGNILQ QFLIEAVLIC VIGGLVGVGL SAAVSLVFNH FVTDFPMDIS
351 AMSVIGAVAC STGIGIAFGF MPANKAAKLN PIDALAQD+
```

35 Computer analysis of this amino acid sequence gave the following results:

Homology with the hypothetical protein o648 of *E. coli* (accession number AE000189) ORF134 and o648 protein show 45% aa identity in 153aa overlap:

```
Orf134: 2
                      RHGTEDFFMNNSDXIRQIVESTTGTMKXXXXXXXXXXXVVGGIGVMNIMLVSVTERTKEI 61
                      RHG +DFF N D + + VE TT T++
                                                             VVGGIGVMNIMLVSVTERT+EI
40
          0648:
                   496 RHGKKDFFTWNMDGVLKTVEKTTRTLQLFLTLVAVISLVVGGIGVMNIMLVSVTERTREI 555
          Orf134: 62
                      GIRMAIGARRGNIXQQFLIEAXXXXXXXXXXXXXXXXXXXXXXXXINHFVTDFPMDISAMSVI 121
                      GIRMA+GAR ++ QQFLIEA
                                                                    F+ + + S ++++
          0648:
                  556 GIRMAVGARASDVLQQFLIEAVLVCLVGGALGITLSLLIAFTLQLFLPGWEIGFSPLALL 615
45
          Orf134: 122 GAVACSTGIGIAFGFMPANKAAKLNPIDALAQD 154
                       A CST GI FG++PA AA+L+P+DALA++
          0648:
                   616 LAFLCSTVTGILFGWLPARNAARLDPVDALARE 648
```

50 Homology with a predicted ORF from N. meningitidis (strain A)

ORF134 shows 98.7% identity over a 154aa overlap with an ORF (ORF134a) from strain A of N. meningi:idis:

							10	•	20	30
	orf134.pep					ARHO	Stedffmni	NSDXIRC	IVESTT(	GTMKLL
22	·								HHHH	111111
	orf134a	GESHTNS	ITVKIK	DNANTQV!	EKGLTD	LLKARHO	STEDFFMNN	ISDSIRC	IVESTT	GTMKLL
		210		220	230		240	250		60
			40	50		60	70		80	90

-313-

		ISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMAIGARRGNIXQQFLIEAVLICVIGG
5		270 280 290 300 310 320
	orf134.pep	100 110 120 130 140 150 LVGVGLSAAVSLVFNHFVTDFPMDISAMSVIGAVACSTGIGIAFGFMPANKAAKLNPIDA
10	orf134a	LVGVGLSAAVSLVFNHFVTDFPMDISAMSVIGAVACSTGIGIAFGFMPANKAAKLNPIDA 330 340 350 360 370 380
		LAQDX
15	orf134a	LAQDX
	The complete length	ORF134a nucleotide sequence <seq 535="" id=""> is:</seq>
		TCGGTGC AAGCAGTATT GGCGCACAAA ATGCGTTCGC TTCTGACGAT
		CGGCATC ATCATCGGTA TCGCTTCGGT TGTCTCCGTC GTCGCATTGG ACGGTTC GCAGAAAAAA ATCCTTGAAG ACATCAGTTC GATAGGGACG
20		ACCATCA GCATCTTCCC AGGGCGCGGC TTCGGCGACA GGCGCAGCGG GATTAAA ACCCTGACCA TAGACGACGC AAAAATCATC GCCAAACAAA
	251 GCT	ACGTTGC TTCCGCCACG CCCATGACTT CGAGCGGCGG CACGCTGACT
		CGCAATA CCGACCTGAC CGCTTCTTTG TACGGTGTGG GCGAACAATA CGACGTG CGCGGGCTGA AGCTGGAAAC GGGGCGGCTG TTTGACGAAA
25	401 ACG	ATGTGAA AGAAGACGCG CAGGTCGTCG TCATCGACCA AAATGTCAAA
		AAACTCT TTGCGGACTC GGATCCGTTG GGTAAAACCA TTTTGTTCAG ACGCCCC TTGACCGTCA TCGGCGTGAT GAAAAAAGAC GAAAACGCTT
		GCAATTC CGACGTGCTG ATGCTTTGGT CGCCCTATAC GACGGTGATG
30		CAAATCA CAGGCGAGAG CCACACCAAC TCCATCACCG TCAAAATCAA CAATGCC AATACCCAGG TTGCCGAAAA AGGGCTGACC GATCTGCTCA
50	701 AAG	CGCGGCA CGGCACGGAA GATTTCTTCA TGAACAACAG CGACAGCATC
		CAGATAG TCGAAAGCAC CACCGGTACG ATGAAGCTGC TGATTTCCTC
		CGCCCTG ATTTCATTGG TAGTCGGCGG CATCGGCGTG ATGAACATCA TGGTGTC CGTTACCGAG CGCACCAAAG AAATCGGCAT ACGGATGGCA
35	901 ATC	GGCGCGC GGCGCGCAA TATTTTGCAG CAGTTTTTGA TTGAGGCGGT
		AATCTGC GTCATCGGCG GTTTGGTCGG CGTGGGTTTG TCCGCCGCCG GCCTCGT GTTCAATCAT TTTGTAACCG ACTTCCCGAT GGACATTTCC
		ATGTCCG TCATCGGCGC GGTCGCCTGT TCGACCGGAA TCGGCATCGC
40		CGGCTTT ATGCCTGCCA ATAAAGCAGC CAAACTCAAT CCGATAGATG
40		TGGCGCA GGATTGA
	Inis encodes a prote	in having amino acid sequence <seq 536="" id="">:</seq>
		QAVLAHK MRSLLTMLGI IIGIASVVSV VALGNGSQKK ILEDISSIGT
		SIFPGRG FGDRRSGRIK TLTIDDAKII AKQSYVASAT PMTSSGGTLT TDLTASL YGVGEQYFDV RGLKLETGRL FDENDVKEDA QVVVIDQNVK
45		FADSDPL GKTILFRKRP LTVIGVMKKD ENAFGNSDVL MLWSPYTTVM
		TGESHTN SITVKIKDNA NTQVAEKGLT DLLKARHGTE DFFMNNSDSI
		VESTTGT MKL <u>LISSIAL ISLVVGGIGV</u> MNIMLVSVTE RTKEIGIRMA RRGNILQ QFLIEAVLIC VIGGLVGVGL SAAVSLVFNH FVTDFPMDIS
		VIGAVAC STGIGIAFGF MPANKAAKLN PIDALAQD*
50	ORF134a and ORF1	34-1 show 100.0% identity in 388 aa overlap:
	orf134a.pep	MSVQAVLAHKMRSLLTMLGIIIGIASVVSVVALGNGSQKKILEDISSIGTNTISIFPGRG
	orf134-1	
55	orf134a.pep	FGDRRSGRIKTLTIDDAKIIAKQSYVASATPMTSSGGTLTYRNTDLTASLYGVGEOYFDV
	orf134-1	
	orf134a.pep	RGLKLETGRLFDENDVKEDAQVVVIDQNVKDKLFADSDPLGKTILFRKRPLTVIGVMKKD
60	orf134-1	
		RGLKLETGRLFDENDVKEDAQVVVIDQNVKDKLFADSDPLGKTILFRKRPLTVIGVMKKD
C F	orf134a.pep	ENAFGNSDVLMLWSPYTTVMHQITGESHTNSITVKIKDNANTQVAEKGLTDLLKARHGTE
65	orf134-1	ENAFGNSDVLMLWSPYTTVMHQITGESHTNSITVKIKDNANTQVAEKGLTDLLKARHGTE

	orf134a.pep	DFFMNNSDSIRQIVESTTGTMKLLISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMA
5	orf134-1	DFFMNNSDSIRQIVESTTGTMKLLISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMA
3	orf134a.pep	IGARRGNILQQFLIEAVLICVIGGLVGVGLSAAVSLVFNHFVTDFPMDISAMSVIGAVAC
•	orf134-1	IGARRGNILQQFLIEAVLICVIGGLVGVGLSAAVSLVFNHFVTDFPMDISAMSVIGAVAC
10	orf134a.pep	STGIGIAFGFMPANKAAKLNPIDALAQDX
	orf134-1	STGIGIAFGFMPANKAAKLNPIDALAQDX

### Homology with a predicted ORF from N.gonorrhoeae

ORF134 shows 96.8% identity over a 154aa overlap with a predicted ORF (ORF134.ng) from N. gonorrhoeae:

•	orf134.pep	ARHGTEDFFMNNSDXIRQIVESTTGTMKLL	30
20	orf134ng		264
	orf134.pep	ISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMAIGARRGNIXQQFLIEAVLICVIGG	90
	orf134ng	ISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMAIGARRGNILQQFLIEAVLICIIGG	324
25	orf134.pep	LVGVGLSAAVSLVFNHFVTDFPMDISAMSVIGAVACSTGIGIAFGFMPANKAAKLNPIDA	150
	orf134ng	LVGVGLSAAVSLVFNHFVTDFPMDISAASVIGAVACSTGIGIAFGFMPANKAAKLNPIDA	384
30	orf134.pep	LAQD 154	İ
	orf134ng	LAQD 388	

## The complete length ORF134ng nucleotide sequence <SEQ ID 537> is:

	1	ATGTCGGTGC	AAGCAGTATT	GGCGCACAAA	ATGCGTTCGC	TTCTGACCAT
	51	GCTCGGCATC	ATCATCGGTA	TCGCTTCGGT	TGTCTCCGTC	GTCGCGCTGG
35	101	GCAACGGTTC	GCAGAAAAA	ATCCTCGAAG	ACATCAGTTC	GATGGGGACG
	151	AACACCATCA	GCATCTTCCC	CGGGCGCGC	TTCGGCGACA	GGCGCAGCGG
	201				AAAAATCATC	
	251	GCTACGTTGC	CTCCGCCACG	CCCATGACTT	CGAGCGGCGG	CACGCTGACC
40	301				TACGGTGTGG	
40	351				GGGGCGGCTG	
	401				TCATCGACCA	
	451				GGTAAAACCA	
	501				GAAAAAAGAC	
45	551				CGCCCTATAC	
45	601				TCCATCACCG	
	651				AGGGCTGGCC	
	701				TGAACAACAG	
	751	AGGCAGATGG			ATGAAGCTGC	
50	801				CATCGGTGTG	
50	851				AAATCGGCAT	
	901				CAGTTTTTGA	
	951				CGTAGGTTTG	TCCGCCGCCG
	1001	TCAGCCTCGT	GTTCAATCAT		ATTTCCCGAT	
	1051	GCGGCATCCG	TTATCGGGGC	GGTCGCCTGT	TCGACCGGAA	TCGGCATCGC
55	1101	GTTCGGCTTT	ATGCCTGCCA	ATAAGGCAGC	CAAACTCAAT	CCGATAGATG
	1151	CATTGGCGCA	GGATTGA	•	•	

### This encodes a protein having amino acid sequence <SEQ ID 538>:

	1	MSVQAVLAHK	MRSLLTMLGI	IIGIASVVSV	VALGNGSQKK	ILEDISSMGT
	51	NTISIFPGRG	FGDRRSGKIK	TLTIDDAKII	AKQSYVASAT	PMTSSGGTLT
60	101	YRNTDLTASL				
	151	DKLFADSDPL	GKTILFRKRP	LTVIGVMKKD	ENAFGNSDVL	MLWSPYTTVM
	201					DFFMNNSDSI
	251	ROMVESTTGT	MKLLISSIAL	<b>ISLVVGGIGV</b>	MNIMLVSVTE	RTKEIGIRMA
•	301					<b>FVTDFPMDIS</b>

### 351 AASVIGAVAC STGIGIAFGF MPANKAAKLN PIDALAQD\*

### ORF134ng and ORF134-1 show 97.9% identity in 388 aa overlap:

5	orf134ng orf134-1	MSVQAVLAHKMRSLLTMLGIIIGIASVVSVVALGNGSQKKILEDISSMGTNTISIFPGRG 
	orf134ng	FGDRRSGKIKTLTIDDAKIIAKQSYVASATPMTSSGGTLTYRNTDLTASLYGVGEQYFDV       :
10	orf134ng	RGLKLETGRLFDENDVKEDAQVVVIDQNVKDKLFADSDPLGKTILFRKRPLTVIGVMKKD
15	orf134ng	ENAFGNSDVLMLWSPYTTVMHQITGESHTNSITVKIKDNANTRVAEKGLAELLKARHGTE
20	orf134-1 orf134ng	ENAFGNSDVLMLWSPYTTVMHQITGESHTNSITVKIKDNANTQVAEKGLTDLLKARHGTE  DFFMNNSDSIRQMVESTTGTMKLLISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMA
	orf134-1 orf134ng	DFFMNNSDSIRQIVESTTGTMKLLISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMA  IGARRGNILQQFLIEAVLICIIGGLVGVGLSAAVSLVFNHFVTDFPMDISAASVIGAVAC
25	orf134-1 orf134ng	IGARRGNILQQFLIEAVLICVIGGLVGVGLSAAVSLVFNHFVTDFPMDISAMSVIGAVAC  STGIGIAFGFMPANKAAKLNPIDALAQDX
30	orf134-1 ORF134ng also show	STGIGIAFGFMPANKAAKLNPIDALAQDX vs homology to an E.coli ABC transporter:
	sp P75831 YBJZ (AE000189) o64 648	Z_ECOLI HYPOTHETICAL ABC TRANSPORTER ATP-BINDING PROTEIN YBJZ >gi5 48; similar to YBBA_HAEIN SW: P45247 [Escherichia coli] Length =
35	Score = 297 Identities =	bits (753), Expect = 6e-80 162/389 (41%), Positives = 230/389 (58%), Gaps = 1/389 (0%)
40	M+	VQAVLAHKMRSLLTMLXXXXXXXXXXXXXXLGNGSQKKILEDISSMGTNTISIFPGRG 60 +A+ A+KMR+LLTML +G+ +++ +L DI S+GTNTI ++PG+ WRALAANKMRTLLTMLGIIIGIASVVSIVVVGDAAKQMVLADIRSIGTNTIDVYPGKD 319
40	FGI	DRRSGKIKTLTIDDAKIIAKQSYVASATPMTSSGGTLTYRNTDLTASLYGVGEQYFDV 120 D + L DD I KQ +VASATP S L Y N D+ AS GV YF+V DDDPQYQQALKYDDLIAIQKQPWVASATPAVSQNLRLRYNNVDVAASANGVSGDYFNV 379
45	G-	LKLETGRLFDENDVKEDAQVVVIDQNVKDKLFAD-SDPLGKTILFRKRPLTVIGVMKK 179 + G F++ + AQVVV+D N + +LF +D +G+ IL P VIGV ++ MTFSEGNTFNQEQLNGRAQVVVLDSNTRRQLFPHKADVVGEVILVGNMPARVIGVAEE 439
50	Query: 180 DEN	NAFGNSDVLMLWSPYTTVMHQITGESHTNSITVKIKDNANTRVAEKGLAELLKARHGT 239 + FG+S VL +W PY+T+ ++ G+S NSITV++K+ ++ AE+ L LL RHG SMFGSSKVLRVWLPYSTMSGRVMGQSWLNSITVRVKEGFDSAEAEQQLTRLLSLRHGK 499
5.5	Query: 240 EDI +DI	FFMNNSDSIRQMVESTTGTMKXXXXXXXXXXVVGGIGVMNIMLVSVTERTKEIGIRM 299 FF N D + + VE TT T++ VVGGIGVMNIMLVSVTERT+EIGIRM
55	Query: 300 AIG	FFTWNMDGVLKTVEKTTRTLQLFLTLVAVISLVVGGIGVMNIMLVSVTERTREIGIRM 559  GARRGNILQQFLIEXXXXXXXXXXXXXXXXXXXXXFNHFVTDFPMDISAASVIGAVA 359  GAR ++LQOFLIE F+ + + S +++ A
60	Sbjct: 560 AV	GARASDVLQQFLIEAVLVCLVGGALGITLSLLIAFTLQLFLPGWEIGFSPLALLLAFL 619 TGIGIAFGFMPANKAAKLNPIDALAQD 388
		T GI FG++PA AA+L+P+DALA++ TVTGILFGWLPARNAARLDPVDALARE 648

Based on this analysis, including the presence of the leader peptide and transmembrane regions in the gonococcal protein, it is prediceted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 65

5 The following partial DNA sequence was identified in N.meningitidis <SEQ ID 539>:

```
..GGGACGGGAG CGATGCTGCT GCTGTTTTAC GCGGTAACGA T.CTGCCTTT
                 51
                       GGCCACTGGC GTTACCCTGA GTTACACCTC GTCGATTTTT TTGGCGGTAT
                101
                       TTTCCTTCCT GATTTTGAAA GAACGGATTT CCGTTTACAC GCAGGCGGTG
                151
                       CTGCTCCTTG GTTTTGCCGG CGTGGTATTG CTGCTTAATC CCTCGTTCCG
10
                201
                       CAGCGGTCAG GAAACGGCGG CACTCGCCGG GCTGGCGGGC GGCGCGATGT
                251
                       CCGGCTGGGC GTATTTGAAA GTGCGCGAAC TGTCTTTGGC GGGCGAACCC
                       GGCTGGCGCG TCGTGTTTTA CCTTTCCGTG ACAGGTGTGG CGATGTCGTC
                301
                351
                       GGTTTGGGCG ACGCTGACCG GCTGGCACAC CCTGTCCTTT CCATCGGCAG
                       TTTATCTGTC GTGCATCGGC GTGTCCGCGC TGATTGCCCA ACTGTCGATG
                401
15
                451
                       ACGCGCGCCT ACAAAGTCGG CGACAAATTC ACGGTTGCCT CGCTTTCCTA
                501
                       TATGACCGTC GTTTTTCCG CTCTGTCTGC CGCATTTTTT CTGGGCGAAG
                       AGCTTTTCTG GCAGGAAATA CTCGGTATGT GCATCATCAT CCTCAGCGGT
                551
                601
                       ATTTTGA
```

This corresponds to the amino acid sequence <SEQ ID 540; ORF135>:

```
20 1 ...GTGAMLLFY AVTILPLATG VTLSYTSSIF LAVFSFLILK ERISVYTQAV
51 LLLGFAGVVL LLNPSFRSQQ ETAALAGLAG GAMSGWAYLK VRELSLAGEP
101 GWRVVFYLSV TGVAMSSVWA TLTGWHTLSF PSAVYLSCIG VSALIAQLSM
151 TRAYKVGDKF TVASLSYMTV VFSALSAAFF LGEELFWQEI LGMCIIISAV
201 F*
```

25 Further work revealed the complete nucleotide sequence <SEQ ID 541>:

```
ATGGATACCG CAAAAAAAGA CATTTTAGGA TCGGGCTGGA TGCTGGTGGC
                51
                    GGCGGCCTGC TTTACCATTA TGAACGTATT GATTAAAGAG GCATCGGCAA
                    AATTTGCCCT CGGCAGCGGC GAATTGGTCT TTTGGCGCAT GCTGTTTTCA
               101
               151
                    ACCGTTGCGC TCGGGGCTGC CGCCGTATTG CGTCGGGACA mCTTCCGCAC
30
                    GCCCCATTGG AAAAACCACT TAAACCGCAG TATGGTCGGG ACGGGGGCGA
               201
               251
                    TGCTGCTGCT GTTTTACGCG GTAACGCATC TGCCTTTGGC CACTGGCGTT
               301
                    ACCCTGAGTT ACACCTCGTC GATTTTTTTG GCGGTATTTT CCTTCCTGAT
               351
                    TTTGAAAGAA CGGATTTCCG TTTACACGCA GGCGGTGCTG CTCCTTGGTT
               401
                    TTGCCGGCGT GGTATTGCTG CTTAATCCCT CGTTCCGCAG CGGTCAGGAA
35
               451
                    ACGGCGGCAC TCGCCGGGCT GGCGGGCGGC GCGATGTCCG GCTGGGCGTA
               501
                    TTTGAAAGTG CGCGAACTGT CTTTGGCGGG CGAACCCGGC TGGCGCGTCG
                    TGTTTTACCT TTCCGTGACA GGTGTGGCGA TGTCGTCGGT TTGGGCGACG
               551
               601
                    CTGACCGGCT GGCACACCCT GTCCTTTCCA TCGGCAGTTT ATCTGTCGTG
                    CATCGGCGTG TCCGCGCTGA TTGCCCAACT GTCGATGACG CGCGCCTACA
               651
40
               701
                    AAGTCGGCGA CAAATTCACG GTTGCCTCGC TTTCCTATAT GACCGTCGTT
                    TTTTCCGCTC TGTCTGCCGC ATTTTTTCTG GGCGAAGAGC TTTTCTGGCA
               801
                    GGAAATACTC GGTATGTGCA TCATCATCCT CAGCGGTATT TTGAGCAGCA
                    TCCGCCCCAC TGCCTTCAAA CAGCGGCTGC AATCCCTGTT CCGCCAAAGA
               851
               901
```

This corresponds to the amino acid sequence <SEQ ID 542; ORF135-1>:

```
1 MDTAKKDILG SGWMLVAAAC FTIMNVLIKE ASAKFALGSG ELVFWRMLFS
51 TVALGAAAVL RRDXFRTPHW KNHLNRSMVG TGAMLLLFYA VTHLPLATGV
101 TLSYTSSIFL AVFSFLILKE RISVYTQAVL LLGFAGVVLL LNPSFRSGQE
151 TAALAGLAGG AMSGWAYLKV RELSLAGEPG WRVVFYLSVT GVAMSSVWAT
201 LTGWHTLSFP SAVYLSCIGV SALIAQLSMT RAYKVGDKFT VASLSYMTVV
251 FSALSAAFFL GEELFWQEIL GMCIIILSGI LSSIRPTAFK QRLQSLFRQR
301 *
```

Computer analysis of this amino acid sequence gave the following results:

-317-

# Homology with a predicted ORF from N.meningitidis (strain A)

ORF135 shows 99.0% identity over a 197aa overlap with an ORF (ORF135a) from strain A of N. meningitidis:

	S	
		10 20 30
5	orfl35.pep	GTGAMLLLFYAVTILPLATGVTLSYTSSIF
	orf135a	STVALGAAAVLRRDTFRTPHWKNHLNRSMVGTGAMLLLFYAVTHLPLATGVTLSYTSSIF 50 60 70 80 90 100
		50 60 70 60 90 100
10		40 50 60 70 80 90
10	orf135.pep	
	orf135a	LAVFSFLILKERISVYTQAVLLLGFAGVVLLLNPSFRSGQETAALAGLAGGAMSGWAYLK
		110 120 130 140 150 160
15		100 110 120 130 140 150
	orf135.pep	
	Offiss.pep	
	orf135a	VRELSLAGEPGWRVVFYLSVTGVAMSSVWATLTGWHTLSFPSAVYLSCIGVSALIAQLSM
20	<b>522</b>	170 180 190 200 210 220
		160 170 180 190 200
	orf135.pep	
25	£125n	
23	orf135a	230 240 250 260 270 280
	orf135a	KQRLQSLFRQRX
		290 300
	m1.41	All ODE125 ample tide comments CEO ID 542 in
30	Ine complete ler	igth ORF135a nucleotide sequence <seq 543="" id=""> is:</seq>
	1	ATGGATACCG CAAAAAAAGA CATTTTAGGA TCGGGCTGGA TGCTGGTGGC
		GGCGGCCTGC TTTACCATTA TGAACGTATT GATTAAAGAG GCATCGGCAA
		AATTTGCCCT CGGCAGCGGC GAATTGGTCT TTTGGCGCAT GCTGTTTTCA
		ACCGTTGCGC TCGGGGCTGC CGCCGTATTG CGTCGGGACA CCTTCCGCAC
35		GCCCCATTGG AAAAACCACT TAAACCGCAG TATGGTCGGG ACGGGGGCGA
	251	TGCTGCTGCT GTTTTACGCG GTAACGCATC TGCCTTTGGC CACCGGCGTT
	301	ACCCTGAGTT ACACCTCGTC GATTTTTTTG GCGGTATTTT CCTTCCTGAT
	351 401	TTTGAAAGAA CGGATTTCCG TTTACACGCA GGCGGTGCTG CTCCTTGGTT TTGCCGGCGT GGTATTGCTG CTTAATCCCT CGTTCCGCAG CGGTCAGGAA
40	451	ACGGCGGCAC TCGCCGGGCT GGCGGGCGGC GCGATGTCCG GCTGGGCGTA
40	501	TTTGAAAGTG CGCGAACTGT CTTTGGCGGG CGAACCCGGC TGGCGCGTCG
	551	TGTTTTACCT TTCCGTGACA GGTGTGGCGA TGTCATCGGT TTGGGCGACG
	601	CTGACCGGCT GGCACACCCT GTCCTTTCCA TCGGCAGTTT ATCTGTCGTG
	651	CATCGGCGTG TCCGCGCTGA TTGCCCAACT GTCGATGACG CGCGCCTACA
45	701	AAGTCGGCGA CAAATTCACG GTTGCCTCGC TTTCCTATAT GACCGTCGTT
	751 801	TTTTCCGCTC TGTCTGCCGC ATTTTTCTG GCCGAAGAGC TTTTCTGGCA GGAAATACTC GGTATGTGCA TCATCATCCT CAGCGGTATT TTGAGCAGCA
	851 851	TCCGCCCCAC TGCCTTCAAA CAGCGGCTGC AATCCCTGTT CCGCCAAAGA
	901	TAA
50	This encodes a p	rotein having amino acid sequence <seq 544="" id="">:</seq>
	1	MDTAKKDILG SGWMLVAAAC FTIMNVLIKE ASAKFALGSG ELVFWRMLFS
	51 101	TVALGAAAVL RRDTFRTPHW KNHLNRSMVG TGAMLLLFYA VTHLPLATGV
	101 151	TLSYTSSIFL AVFSFLILKE RISVYTQAVL LLGFAGVVLL LNPSFRSGQE TAALAGLAGG AMSGWAYLKV RELSLAGEPG WRVVFYLSVT GVAMSSVWAT
55	201	LIGHTLSFP SAVYLSCIGV SALIAQLSMI RAYKVGDKFT VASLSYMIVV
55	251	FSALSAAFFL AEELFWQEIL GMCIIILSGI LSSIRPTAFK QRLQSLFRQR
	301	*
	ODE125 1 O	DE125 1 above 00 20/ identity in 200 aventon.
	OKF133a and O	RF135-1 show 99.3% identity in 300 aa overlap:

orf135a.pep

orf135-1

60

 ${\tt MDTAKKDILGSGWMLVAAACFTIMNVLIKEASAKFALGSGELVFWRMLFSTVALGAAAVL}$ 

	orf135a.pep	RRDTFRTPHWKNHLNRSMVGTGAMLLLFYAVTHLPLATGVTLSYTSSIFLAVFSFLILKE
5	orf135-1	RRDXFRTPHWKNHLNRSMVGTGAMLLLFYAVTHLPLATGVTLSYTSSIFLAVFSFLILKE
	orf135a.pep	RISVYTQAVLLLGFAGVVLLLNPSFRSGQETAALAGLAGGAMSGWAYLKVRELSLAGEPG
	orf135-1	RISVYTQAVLLLGFAGVVLLLNPSFRSGQETAALAGLAGGAMSGWAYLKVRELSLAGEPG
10	orf135a.pep	WRVVFYLSVTGVAMSSVWATLTGWHTLSFPSAVYLSCIGVSALIAQLSMTRAYKVGDKFT
	orf135-1	WRVVFYLSVTGVAMSSVWATLTGWHTLSFPSAVYLSCIGVSALIAQLSMTRAYKVGDKFT
15	orf135a.pep	VASLSYMTVVFSALSAAFFLAEELFWQEILGMCIIILSGILSSIRPTAFKQRLQSLFRQR
13	orf135-1	VASLSYMTVVFSALSAAFFLGEELFWQEILGMCIIILSGILSSIRPŢAFKQRLQSLFRQR

### Homology with a predicted ORF from N. gonorrhoeae

ORF135 shows 97% identity over a 201aa overlap with a predicted ORF (ORF135ng) from

### 20 N.gonorrhoeae:

	orf135.pep	GTGAMLLLFYAVTXLPLATGVTLSYTSSIF	30
	orf135ng		335
25	orf135.pep	LAVFSFLILKERISVYTQAVLLLGFAGVVLLLNPSFRSGQETAALAGLAGGAMSGWAYLK	90
	orf135ng	LAVFSFLILKERISVYTQAVLLLGFAGVVLLLNPSFRSGQEPAALAGLAGGAMSGWAYLK	395
30	orf135.pep	VRELSLAGEPGWRVVFYLSVTGVAMSSVWATLTGWHTLSFPSAVYLSCIGVSALIAQLSM	150
	orf135ng	VRELSLAGEPGWRVVFYLSATGVAMSSVWATLTGWHTLSFPSAVYLSGIGVSALIAQLSM	455
	orf135.pep	TRAYKVGDKFTVASLSYMTVVFSALSAAFFLGEELFWQEILGMCIIISAVF 201	
35	orf135ng	TRAYKVGDKFTVASLSYMTVVFSALSAAFFLGEELFWQEILGMCIIISAAF 506	

An ORF135ng nucleotide sequence <SEQ ID 545> was predicted to encode a protein having amino acid sequence <SEQ ID 546>:

```
MPSEKAFRRH LRTASFQGLH LHHFHQKVGK CGIIGFGIHI FPTLLPAAQG
                         ILDIQLGLFR IDFAALAVYR RTQVDFIHTV IDGIASDQAF SEVVQILRRL
40
                        NLGHFTDTHL IAQARRFIAD FGNIRPMRRG EAKTFCRCFR FDGIDGIHGD FRQCGHINRL APGKDCRNGK RDKVFFHTRH YNQVCLEKTN CSARKIKFRH
                   101
                   151
                   201
                         QKQAKTHSTS LAARFTIRPS LSQRPFMDTA KKDILGSGWM LVAAACFTVM
                   251
                         NVLIKEASAK FALGSGELVF WRMLFSTVTL GAAAVLRRDT FRTPHWKNHL
                         NRSMVGTGAM LLLFYAVTHL PLTTGVTLSY TSSIFLAVFS FLILKERISV
YTQAVLLLGF AGVVLLLNPS FRSGQEPAAL AGLAGGAMSG WAYLKVRELS
                   301
45
                   351
                         LAGEPGWRVV FYLSATGVAM SSVWATLTGW HTLSFPSAVY LSGIGVSALI
                   401
                   451
                         AQLSMTRAYK VGDKFTVASL SYMTVVFSAL SAAFFLGEEL FWQEILGMCI
                   501
                         IISAAF*
```

#### Further work revealed the following gonococcal sequence <SEQ ID 547>:

•						
50	1	ATGGATACCG	CAAAAAAAGA	CATTTTAGGA	TCGGGCTGGA	TGCTGGTGGC
•	51	GGCGGCCTGC	TTCACCGTTA	TGAACGTATT	GATTAAAGAG	GCATCGGCAA
	101	AATTTGCCCT	CGGCAGCGGC	GAATTGGTCT	TTTGGCGCAT	GCTGTTTTCA
	151	ACCGTTACGC	TCGGTGCTGC	CGCCGTATTG	CGGCGCGACA	CCTTCCGCAC
	201	GCCCCATTGG	AAAAACCACT	TAAACCGCAG	TATGGTCGGG	ACGGGGGCGA
55	251	TGCTGCTGCT	<b>GTTTTACGCG</b>	GTAACGCATC	TGCCTTTGAC	AACCGGCGTT
•	301	ACCCTGAGTT	ACACCTCGTC	GATTTTTttg	GCGGTATTTT	CCTTCCTGAT
	351	TTTGAAAGAA	CGGATTTCCG	TTTACACGCA	GGCGGTGCTG	CTCCTTGGTT
	401	TTGCCGGCGT	GGTATTGCTG	CTTAATCCCT	CGTTCCGCAG	CGGTCAGGAA
	451	CCGGCGGCAC	TCGCCGGGCT	GGCGGGCGGC	GCGATGTCCG	GCTGGGCGTA
60	501	TTTGAAAGTG	CGCGAACTGT	CTTTGGCGGG	CGAACCCGGC	TGGCGCGTCG
	551	TGTTTTACCT	TTCCGCAACC	GGCGTGGCGA	TGTCGTCggt	ttgggcgacg
	601	Ctgaccggct	ggCACAcccT	GTCCTTTcca	tcggcagttt.	ATCtgtCGGG

```
651 CATCGGCGTG tccgcgCtgA TTGCCCAaCT GtcgatgAcg cGCGcctaca
701 aaGTCGGCGA CAAATTCACG GTTGCCTCGC tttcctaTAt gaccgtcGTC
751 TTTTCCGCCC TGTCTGCCGC ATTTTTTCTG ggcgaagagc tttTCtggCA
801 GGAAATACTC GGTATGTGCA TCATTAtccT CAGCGGCATT TTGAGCAGCA
5 851 TCCGCCCCAT TGCCTTCAAA CAGCGGCTGC AAGCCCTCTT CCGCCAAAGA
901 TAA
```

This corresponds to the amino acid sequence <SEQ ID 548; ORF135ng-1>:

```
10 MDTAKKDILG SGWMLVAAAC FTVMNVLIKE ASAKFALGSG ELVFWRMLFS
51 TVTLGAAAVL RRDTFRTPHW KNHLNRSMVG TGAMLLLFYA VTHLPLTTGV
101 TLSYTSSIFL AVFSFLILKE RISVYTQAVL LLGFAGVVLL LNPSFRSQE
151 PAALAGLAGG AMSGWAYLKV RELSLAGEPG WRVVFYLSAT GVAMSSVWAT
201 LTGWHTLSFP SAVYLSGIGV SALIAQLSMT RAYKVGDKFT VASLSYMTVV
251 FSALSAAFFL GEELFWQEIL GMCIIILSGI LSSIRPIAFK QRLQALFRQR
301 *
```

ORF135ng-1 and ORF135-1 show 97.0% identity in 300 aa overlap:

```
orf135ng-1.pep MDTAKKDILGSGWMLVAAACFTVMNVLIKEASAKFALGSGELVFWRMLFSTVTLGAAAVL
                    MDTAKKDILGSGWMLVAAACFTIMNVLIKEASAKFALGSGELVFWRMLFSTVALGAAAVL
        orf135-1
        orf135ng-1.pep RRDTFRTPHWKNHLNRSMVGTGAMLLLFYAVTHLPLTTGVTLSYTSSIFLAVFSFLILKE
20
                    RRDXFRTPHWKNHLNRSMVGTGAMLLLFYAVTHLPLATGVTLSYTSSIFLAVFSFLILKE
        orf135-1
        orf135ng-1.pep RISVYTQAVLLLGFAGVVLLLNPSFRSGQEPAALAGLAGGAMSGWAYLKVRELSLAGEPG
                    25
                    RISVYTQAVLLLGFAGVVLLLNPSFRSGQETAALAGLAGGAMSGWAYLKVRELSLAGEPG
        orf135-1
        orf135ng-1.pep WRVVFYLSATGVAMSSVWATLTGWHTLSFPSAVYLSGIGVSALIAQLSMTRAYKVGDKFT
                    пини:ининининининин инининини
                    WRVVFYLSVTGVAMSSVWATLTGWHTLSFPSAVYLSCIGVSALIAQLSMTRAYKVGDKFT
30
        orf135-1
        orf135ng-1.pep VASLSYMTVVFSALSAAFFLGEELFWQEILGMCIIILSGILSSIRPIAFKQRLQALFRQR
                    ությունում անտանական արտարական արտարական
                    VASLSYMTVVFSALSAAFFLGEELFWQEILGMCIIILSGILSSIRPTAFKQRLQSLFRQR
         orf135-1
```

Based on this analysis, including the presence of several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 66

The following DNA sequence was identified in N.meningitidis <SEQ ID 549>:

40	1 51	T	CTGTTGCCGA	AAATCGTCAA	TTCCCGCAGA TACAGTTCCG TCTTCATACA	GCACATCGGA
	101 151 201	CTGCCCGGGA	TCGCCGAAAT	CCGCCCATTC	TGCGGCATCG	AAAGCCGCCG
45	251 301	TAGGGGATGC	CTTTCGCCTT	GAACATCCAG GTTCGACATT	TCGCTGATGT GGTCAGTTCG	CGTCAACCGG
	351 401	САСАТАТСТТ	CGCAAATTTC	GCCGTCTTCG	CAAAATAAAT CCGTCTTGGA	AAAAAGGGAC
50	451 501 551	AAAGcTCGCG	CCAAAAATAT	TTGAATGTTT	AATGCGGCGG TACGGGCGCG TAATAAATGA	CGGAATCGCC
	601 651 701	CATCATATCT GCTTTCTgcC GAATAG	GCTCCTCAAC kTCGGCATCC	GTGTACGGTA GATTCGGATT	TCTGTTTGCA TGAAAAGTTC	CCTTACTGCG mmrwyATTCG

55 This corresponds to the amino acid sequence <SEQ ID 550; ORF136>:

<sup>1</sup> MKRRIAVFVL FPQIIRVLGQ LLPKIVNTVP AHRMLFQIFG MFFFFIHQQY 51 LPGIAEIDSP CGIVFGALLF RHLPAHCLYG KAAVGDAVAH EHPVADVVNR

5

10

15

20

25

30

35

40

45

50

55

201 251

```
NANAFALFDI GQFAXFIVQH TVNIKTVKIN IVDPHMFANF AVFAVLEKRD
         101
               FDHGKIQGGN NAAAFPKKLA PKIFECFTGA FVGTVYRFVC LFYIINDGIA
              HHSAPQRVRY LFAPYCGFLP SASDSDLKSS XXSE*
         201
Further work revealed the complete nucleotide sequence <SEQ ID 551>:
              ATGATGAAGC GGCGTATAGC CGTCTTCGTC CTGTTCCCGC AGATAATCCG
           51
               AGTTTTGGGA CAACTGTTGC CGAAAATCGT CAATACAGTT CCGGCACATC
              GGATGCTCTT CCAGATTTTC GGGATGTTCT TTTTCTTCAT ACACCAGCAA
         101
               TATCTGCCCG GGATCGCCGA AATCGATTCC CCATGCGGCA TCGTGTTCGG
         151
              TGCGCTCCTC TTCCGTCATC TGCCCGCGCA TTGCCTGTAT GGTAAAGCCG
          201
               CCGTAGGGGA TGCCGTTGCA CACGAACATC CAGTCGCTGA TGTCGTCAAC
          251
               CGGAACGCAA ACGCTTTCGC CTTGTTCGAC ATTGGTCAGT TCGCCGGGTT
          301
               CATTGTTCAG CACACCGTAA ATATAAAGAC CGTCAAAATA AATATCGTCG
          351
          401
              ATCCACATAT GTTCGCAAAT TTCGCCGTCT TCGCCGTCTT GGAAAAAAGG
          451
               GACTTTGACC ATGGCAAAAT CCAAGGCGGA AATAATGCGG CGGCGTTCCC
              AAAAAAGCTC GCGCCAAAAA TATTTGAATG TTTTACGGGC GCGTTCGTCG
          501
              GCACGGTTTA CCGGTTCGTC TGCCTGTTCT ACATAATAAA TGACGGAATC
          551
          601
               GCCCATCATT CTGCTCCTCA ACGTGTACGG TATCTGTTTG CACCTTACTG
               CGGCTTTCTG CCTTCGGCAT CCGATTCGGA TTTGAAAAGT TCCAAATATT
          651
          701
               CGGAATAG
This corresponds to the amino acid sequence <SEQ ID 552; ORF136-1>:
               MMKRRIAVFV LFPQIIRVLG QLLPKIVNTV PAHRMLFQIF GMFFFFIHQQ
           51
               YLPGIAEIDS PCGIVFGALL FRHLPAHCLY GKAAVGDAVA HEHPVADVVN
          101
               RNANAFALFD IGQFAGFIVQ HTVNIKTVKI NIVDPHMFAN FAVFAVLEKR
               DFDHGKIQGG NNAAAFPKKL APKIFECFTG AFVGTVYRFV CLFYIINDGI
               AHHSAPQRVR YLFAPYCGFL PSASDSDLKS SKYSE*
Computer analysis of this amino acid sequence gave the following results:
Homology with a predicted ORF from N. meningitidis (strain A)
ORF136 shows 71.7% identity over a 237aa overlap with an ORF (ORF136a) from strain A of N.
meningitidis:
                                              30
     orf136.pep
                   MKRRIAVFVLFPQIIRVLGQLLPKIVNTVPAHRMLFQIFGMFFFFIHQQYLPGIAEIDS
                   orf136a
                  MMKRRIAVFVLLMQKIRILGQLLPKIVNTVPAHRMLFQXFGMFFFFIHQQYLPGIAEIDS
                         10
                                   20
                                                      40
                                             30
                                                                50
                 60
                          70
                                    80
                                              90
                                                      100
                                                                110
                                                                         119
     orf136.pep
                  PCGIVFGALLFRHLPAHCLYGKAAVGDAVAHEHPVADVVNRNANAFALFDIGQFAXFIVQ
                  orf136a
                  PCGIVFGTLLFRHXSTHCLYGKAAVGNAVAHEHPVADVVNRNANAFALFDIGOFAGFIVO
                         70
                                   80
                                             90
                                                     100
                                                               110
                120
                                   140
                          130
                                             150
                                                      160
                                                                170
     orf136.pep
                  HTVNIKTVKINIVDPHMFANFAVFAVLEKRDFDHGKIQGGNNAAAFPKKLAPKIFECFTG
                  orf136a
                  HAINVKTVKINIVDPHMFANFAXFAVLEKRALTMAKSKXXXMRRRSQKSSRQKYLNVLRA
                         130
                                  140
                                            150
                                                     160
                                                               170
                180
                          190
                                   200
                                                210
                                                          220
     orf136.pep
                  AFVGTVYRFVCLFYIINDGIAHH---SAPQRVRYLFAPYCGFLPSASDSDLKSSXXSEX
                        11: 1
                                    :::
                                           111111111111111111111111111111
                  R---SPARFTGLSACSTXXMTESPIISAPQRVRYLFAPYCGFLPSASDSDLKSSKYSEX
     orf136a
                            190
                                     200
 The complete length ORF136a nucleotide sequence <SEQ ID 553> is:
               ATGATGAAGC GGCGTATAGC CGTCTTCGTC CTGCTCATGC AGAAAATCCG
               GATTTTGGGA CAACTGTTGC CGAAAATCGT CAATACAGTT CCGGCACATC
          101
               GGATGCTCTT CCAGATNTTC GGGATGTTCT TTTTCTTCAT ACACCAGCAA
```

TACCTGCCCG GGATCGCCGA AATCGATTCC CCATGCGGCA TCGTGTTCGG TACGCTCCTC TTCCGTCATC NGTCCACGCA TTGCCTGTAT GGTAAAGCCG

CCGTAGGGAA TGCCGTTGCA CACGAACATC CAGTCGCTGA TGTCGTCAAC

5 10	351 CAT 401 ATC 451 GCT 501 AAA 551 CAC 601 CCC 651 CGC 701 CGC This encodes a prote  1 MMF 51 YLE 101 RNA 151 ALT	AACGCAA ACGCTTTCGC CTTGTTCGAC ATTGGTCAGT TCGCCGGGTT TGTTCAG CACGCCATAA ATGTAAAGAC CGTCAAAATA AATATCGTCG CACACATAT GTTCGCAAAT TTCGCCNTCT TCGCCGTCTT GGAAAAAAGG TTTGACCA TGGCAAAAT ATTTGAATGT TTTGCGGGCG GGCGTTCCCA CAGACTCG CGCCAAAAAT ATTTGAATGT TTTGCGGGCG GGCGTTCCCCG CGTTTAC CGGTTTGTCT GCCTGTTCTA CATAATAAAT GACGGAATCG CACCTTACTG CCTTCCGCA CCGATTCGGA TTTTGAAAAGT TCCAAATATT CAATAG  CIN having amino acid sequence <seq 554="" id="">:  CRRIAVFV LLMQKIRILG QLLPKIVNTV PAHRMLFQXF GMFFFFIHQQ CGIAEIDS PCGIVFGTLL FRHXSTHCLY GKAAVGNAVA HEHPVADVVN ANAFALFD IGQFAGFIVQ HAINVKTVKI NIVDPHMFAN FAXFAVLEKR CMAKSKXX XMRRSQKSS RQKYLNVLRA RSPARFTGLS ACST**MTES</seq>	
15		SAPORVR YLFAPYCGFL PSASDSDLKS SKYSE*  136-1 show 73.1% identity in 238 aa overlap:	
	014 1004 411 011	10 20 30 40 50 60	
20	orf136a.pep	MMKRRIAVFVLLMQKIRILGQLLPKIVNTVPAHRMLFQXFGMFFFFIHQQYLPGIAEIDS	
25	orf136a.pep	70 80 90 100 110 120 PCGIVFGTLLFRHXSTHCLYGKAAVGNAVAHEHPVADVVNRNANAFALFDIGQFAGFIVQ	
30	orf136a.pep orf136-1	130 140 150 160 170 180 HAINVKTVKINIVDPHMFANFAXFAVLEKRALTMAKSKXXXMRRRSQKSSRQKYLNVLRA  :: :	
35	orf136a.pep orf136-1	190 200 210 220 230  RSPARFTGLSACSTXXMTESPIISAPQRVRYLFAPYCGFLPSASDSDLKSSKYSEX :   :   : :::	
40	Homology with a p	redicted ORF from N.gonorrhoeae	
		3% identity over a 234aa overlap with a predicted ORF (ORF136ng) from	m
	N.gonorrhoeae:	, , , , , , , , , , , , , , , , , , ,	
45	orf136.pep orf136ng	MKRRIAVFVLFPQIIRVLGQLLPKIVNTVPAHRMLFQIFGMFFFFIHQQYLPGIAEIDS 59	
50	orf136.pep orf136ng	PCGIVFGALLFRHLPAHCLYGKAAVGDAVAHEHPVADVVNRNANAFALFDIGQFAXFIVQ 119 	
55	orf136.pep orf136ng	HTVNIKTVKINIVDPHMFANFAVFAVLEKRDFDHGKIQGGNNAAAFPKKLAPKIFECFTG 179	
<i>) )</i>	orf136.pep orf136ng	AFVGTVYRFVCLFYIINDGIAHHSAPQRVRYLFAPYCGFLPSASDSDLKSSXXSE 234   :	

The complete length ORF136ng nucleotide sequence <SEQ ID 555> is:

60

<sup>1</sup> ATGATGAAGC GGCGTATAGC CGTCTTCGTC CTGCTCATGC AGAAAATCCG 51 GATTTTGGGA CAACTGTTGC CGAAAATCGT CAATACAGTT CCGGCACATC

	101	GGATGCTCTT	CCAAATTTTC	GGGATGTTCT	TTTTCTTCAT	ACACCGGCAA
	151	TACCTGCCCG	GGATCGCCGA	AATCGATTCC	CCAGGCGGTA	TCGTGTTCGG
	201	TACGCTCCTC	TTCCGTCATC	TGTCCGCGCA	TTGCCTGTAC	GGTAAAGCCG
•	251	CCGTAGGGGA	TGCCGTTGCA	CACGAACATC	CAGTCGCTGA	TGTCGCCAAC
5	301	CGGAACGCAA	ACGCTTTCGC	CTTGTTCGAC	ATTGGTCAGT	CCGCCGGGTT
•	351	CATTGTTCAG	CACACCGTAA	ATATAAAGAC	CGTCAAAATA	AATATCGTCG
	401	ATCCACATAT	GTTCGCAAAT	TTCGCCGTCT	TCGCCGTCTT	GGAAAAAAGG
	451	GACTTTGACC	ATGGCAAAAT	CCAAGGCGGA	AATAATGCGG	CGGCGTTCCC
	501	AAAAAAGCTC	GCGCCAAAAG	TATTTGAATG	TTTTACGGGC	GCGTTCGCCG
10	551	GCACGGTTTA	CCGGTTCGTC	TGCCTGTTCT	ACATAATAAA	TGACGGAATC
	601	GCCCATCATA	CTGCTCCTCA	ACGTGTACGG	TATCTGTTTG	CACCTTACCG
	651	CGGTTTTCTA	CCTCCGGCAT	CCGATTCGGA	TTTGAAAAGT	TCCAAATATT
· ·	701	CGGAATAG				

This encodes a protein having amino acid sequence <SEQ ID 556>:

```
1 MMKRRIAVFV LLMQKIRILG QLLPKIVNTV PAHRMLFQIF GMFFFFIHRQ
51 YLPGIAEIDS PGGIVFGTLL FRHLSAHCLY GKAAVGDAVA HEHPVADVAN
101 RNANAFALFD IGQSAGFIVQ HTVNIKTVKI NIVDPHMFAN FAVFAVLEKR
151 DFDHGKIQGG NNAAAFPKKL APKVFECFTG AFAGTVYRFV CLFYIINDGI
201 AHHTAPQRVR YLFAPYRGFL PPASDSDLKS SKYSE*
```

#### 20 ORF136ng and ORF136-1 show 93.6% identity in 235 aa overlap:

	orf136ng	MMKRRIAVFVLLMQKIRILGQLLPKIVNTVPAHRMLFQIFGMFFFFIHRQYLPGIAEIDS
•	orf136-1	MMKRRIAVFVLFPQIIRVLGQLLPKIVNTVPAHRMLFQIFGMFFFFIHQQYLPGIAEIDS
25	orf136ng	PGGIVFGTLLFRHLSAHCLYGKAAVGDAVAHEHPVADVANRNANAFALFDIGQSAGFIVQ
	orf136-1	PCGIVFGALLFRHLPAHCLYGKAAVGDAVAHEHPVADVVNRNANAFALFDIGQFAGFIVQ
30	orf136ng	HTVNIKTVKINIVDPHMFANFAVFAVLEKRDFDHGKIQGGNNAAAFPKKLAPKVFECFTG
	orf136-1	HTVNIKTVKINIVDPHMFANFAVFAVLEKRDFDHGKIQGGNNAAAFPKKLAPKIFECFTG
	orf136ng	AFAGTVYRFVCLFYIINDGIAHHTAPQRVRYLFAPYRGFLPPASDSDLKSSKYSEX
35	orf136-1	AFVGTVYRFVCLFYIINDGIAHHSAPQRVRYLFAPYCGFLPSASDSDLKSSKYSEX

Based on the presence of the putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 67

40 The following partial DNA sequence was identified in N.meningitidis <SEQ ID 557>:

```
ATGGAAAATA TGGTAACGTT TTCAAAAATC AGACCGCTTT TGGCAATCGC
                    CGCCGCCGCG TTGCTTGCCG CC.TGCGGAC GGCGGGAAAT AATGCTGTCC
               101
                    GCAAGCCGGT GCAAACCGCC AAACCCGCCG CAGTGGTCGG TTTGGCACTC
               151
                    GGTGGCGGCG CATCTAAAGG ATTTGCCCAT GTAGGTATTA TTAAGGTTTT
45
               201
                    GAAAGAAAAC GGTATTCCTG TGAAGGTGGT TACCGGCACC TCCGCAGGTT
                    CGATTGTCGG CAACCTTTTT GCATCGGGTA TGTCGCCCGA CCGCCTCGAA
               251
               301
                    TTGGAAGCCG AAATTTTAGG CAAAACCGAT TTGGTCGATT TAACCTTGTC
                    CACCAATGGG TTTATCAAAG GCGCAAAGCT GCAAAATTAC ATCAACCGAA
               351
                    AACTCCGCGG CATGCAGATT CAGCAGTTTC CCATCAAATT TGCCGCC..
```

This corresponds to the amino acid sequence <SEQ ID 558; ORF137>:

```
1 MENMVTFSKI RPLLAIAAAA LLAAXRTAGN NAVRKPVQTA KPAAVVGLAL
51 GGGASKGFAH VGIIKVLKEN GIPVKVVTGT SAGSIVGNLF ASGMSPDRLE
```

101 LEAEILGKTD LVDLTLSTNG FIKGAKLQNY INRKLRGMQI QQFPIKFAA...

Further work revealed the complete nucleotide sequence <SEQ ID 559>:

				-323-		
			•			
	1	ATGGAAAATA	TGGTAACGTT	TTCAAAAATC	AGACCGCTTT	TGGCAATCGC
	51	CCCCCCCCC	TTGCTTGCCG	CCTGCGGCAC	GGCGGGAAAT	AATGCTGTCC
	101	CCAACCCGGT	GCAAACCGCC	AAACCCGCCG	CAGTGGTCGG	TTTGGCACTC
	151	CCTCCCCCCC	CATCTAAAGG	ATTTGCCCAT	GTAGGTATTA	TTAAGGTTTT
5	201	CAAACAAAAC	GGTATTCCTG	TGAAGGTGGT	TACCGGCACA	TCGGCAGGTT
5	251	CCATTCTCGG	CAGCCTTTTT	GCATCGGGTA	TGTCGCCCGA	CCGCCTCGAA
	301	TTCCAACCCG	AAATTTTAGG	CAAAACCGAT	TTGGTCGATT	TAACCTTGTC
	351	CACCACTCCT	TTTATCAAAG	GCGAAAAGCT	GCAAAATTAC	ATCAACCGAA
	401	AACTCGGCGG	CAGGCAGATT	CAGCAGTTTC	CCATCAAATT	TGCCGCCGTT
10	451	ርርጥልርጥርልጥጥ	TTGAAACCGG	CAAGGCCGTC	GCTTTCAATC	AGGGGAATGC
10	501	CCCCCACCCT	GTGCGCGCTT	CCGCCGCCAT	TCCCAATGTG	TTCCAACCCG
	551	TTATCATCGG	CAGGCATACA	TATGTTGACG	GCGGTCTGTC	GCAGCCCGTG
	601	CCCCTCAGTG	CCGCCCGGCG	GCAGGGGGCG	AATTTCGTGA	TTGCCGTCGA
	651	TATTTCCCCC	CGTCCGGGCA	AAAACATCAG	CCAAGGTTTC	TTCTCTTATC
15	701	TOGATOAGAC	GCTGAACGTA	ATGAGCGTTT	CTGCGTTGCA	AAATGAGTTG
13	751	CCCCACCCG	ATGTGGTTAT	CAAACCGCAG	GTTTTGGATT	TGGGTGCAGT
	801	CGGCGGATTC	GATCAGAAAA	AACGCGCCAT	CCGGTTGGGT	GAGGAGGCAG
	851	CACGTGCCGC	ATTGCCTGAA	ATCAAACGCA	AACTGGCGGC	ATACCGTTAT
	901	TGA				
				-OFO F	D SCO. ODE	127 1~.
20	This correspond	is to the amu	no acid seque	ence <seq l<="" th=""><th>D 300; UKF</th><th>13/-1/:</th></seq>	D 300; UKF	13/-1/:
	•					
	1	MENMVTFSKI	RPLLAIAAA	LLAACGTAGN	NAVRKPVQTA	KPAAVVGLAL
	51	GGGASKGFAH	VGIIKVLKEN	GIPVKVVTGT	SAGSIVGSLF	ASGMSPDRLE
	101	LEAEILGKTD	LVDLTLSTS	FIKGEKLQNY	INRKVGGRQI	QQFPIKFAAV
	151	ATDFETGKAV	AFNQGNAGQA	VRASAAIPNV	FQPVIIGRHT	YVDGGLSQPV
25	201	PVSAARRQGA	NFVIAVDISA	RPGKNISQGF	· FSYLDQTLNV	MSVSALQNEL
20	251	GQADVVIKPO	VLDLGAVGGI	DOKKRAIRLO	EEAARAALPE	IKRKLAAYRY
	301	*				
		C.II		~~~~	the following	a reculte.
	Computer analy	sis of this at	mino acid se	quence gave	THE TOTTOWITE	g resures.
	-					
		4 1	ODE Com A	T	a (atrain A)	

# Homology with a predicted ORF from N. meningitidis (strain A)

ORF137 shows 93.3% identity over a 149aa overlap with an ORF (ORF137a) from strain A of N. meningitidis:

		10	20	30	40	50	60
	orf137.pep	MENMVTFSKIRPLLA	IAAAALLA		KPVQTAKPA	AVVGLALGGGAS	SKGFAH
			 	:    CCM2 CNN/2 2B			
35	orf137a	MENMVTFSKIRPLLA 10	таааацыа 20	30	40	50	60
		10	20	30			
		70	80	90	100	110	120
	orf137.pep	VGIIKVLKENGIPVK	VVTGTSAG	SIVGNLFASGM	SPDRLELEA	ILGKTDLVDL'	TLSTNG
40	• •		11111111	1111:11111			:  TLSTSG
	orf137a	VGIIKVLKENGIPVK	VVTGTSAG 80	SIVGSLFASGM 90	SPURLELEA 100	110	120
		70	80	90	100	110	
		130	140	149			
45	orf137.pep	FIKGAKLQNYINRKI	RGMQIQQF	PIKFAA			
45		- 1111 11111111111111111111111111111111	1:1111	11111		~~\\\\ ~~\\\\\\\\\\\\\\\\\\\\\\\\\\\\\	יוואם ד ה הי
	orf137a	FIKGEKLONYINRK		PIKFAAVATDE 150	ETGKAVAFN 160	QGNAGQAVKAS 170	180
		130	140	120	100	170	100

The complete length ORF137a nucleotide sequence <SEQ ID 561> is:

50 55	1 51 101 151 201 251 301 351 401 451 501	CGCCGCCGC GCAAGCCGGT GGTGGCGCG GAAAGAAAAC CGATAGTCGG TTGGAAGCCG CACCAGTGGT AAGTCGGCGG GCTACTGATT CGGGCAGGCT	TGGTAACGTT TTGCTTGCCG GCAAACCGCC CATCTAAAGG GGTATTCCTG CAGCCTTTTT AAATTTTAGG TTTATCAAAG CAGGCGGATT TTGAAACCGG GTGCGCGCTT	CCTGCGGCAC AAACCCGCCG ATTTGCCCAT TGAAGGTGGT GCATCGGGTA TAAAACCGAT GCGAAAAGCT CAGCAGTTCC CAAGGCCGTC CCGCCGCCAT	GGCGGGAAAT CAGTGGTCGG GTAGGTATTA TACCGGCACA TGTCGCCCGA TTGGTCGATT GCAAAATTAC CCATCAAATT GCTTTCAATC TCCCAATGTG	AATGCTGCCC TTTGGCACTC TTAAGGTTTT TCGGCAGGTT CCGCCTCGAA TAACCTTGTC ATCAACCGAA TGCCGCCGTT AAGGGAATGC TTCCAACCCG
60		CGGGCAGGCT	CAGGCATACA	TATCTTCACG	GCGGTCTGTC	GCAGCCCGTG
	551	TTATCATCGG	CHGGCTTACA	INIGIIGACG	000010101	

5	601 CCCGTCAGTG CCGCCCGGCG GCANGNNNNG NATNTCGTGA TTGCCGTCGA 651 TATTTCCGCC CGTCCGAGCA AAACATCAG CCAAGGCTTC TTCTCTTATC 701 TCGATCAGAC GCTGAACGTA ATGAGCGTTT CCGCGTTGCA AAATGAGTTG 751 GGGCAGGCGG ATGTGGTTAT CAAACCGCAG GTTTTGGATT TGGGTGCAGT 801 CGGCGGATTC GATCAGAAAA AACGCGCCAT CCGGTTGGGT GAGGAGGCAG 851 CACGTGCCGC ATTGCCTGAA ATCAAACGCA AACTGGCGGC ATACCGTTAT 901 TGA	
	This encodes a protein having amino acid sequence <seq 562="" id="">:</seq>	
10	1 MENMYTFSKI RPLLAIAAAA LLAACGTAGN NAARKPVQTA KPAAVVGLAL 51 GGGASKGFAH VGIIKVLKEN GIPVKVVTGT SAGSIVGSLF ASGMSPDRLE 101 LEAEILGKTD LVDLTLSTSG FIKGEKLQNY INRKVGGRRI QQFPIKFAAV 151 ATDFETGKAV AFNQGNAGQA VRASAAIPNV FQPVIIGRHT YVDGGLSQPV 201 PVSAARRXXX XXVIAVDISA RPSKNISQGF FSYLDQTLNV MSVSALQNEL 251 GQADVVIKPQ VLDLGAVGGF DQKKRAIRLG EEAARAALPE IKRKLAAYRY	
15	301 *	
	ORF137a and ORF137-1 show 97.3% identity in 300 aa overlap:	
20	orf137a.pep MENMVTFSKIRPLLAIAAAALLAACGTAGNNAARKPVQTAKPAAVVGLALGGGASKGFAH	
20	orf137a.pep VGIIKVLKENGIPVKVVTGTSAGSIVGSLFASGMSPDRLELEAEILGKTDLVDLTLSTSG	
•	orf137-1 VGIIKVLKENGIPVKVVTGTSAGSIVGSLFASGMSPDRLELEAEILGKTDLVDLTLSTSG	
25	orf137a.pep FIKGEKLQNYINRKVGGRRIQQFPIKFAAVATDFETGKAVAFNQGNAGQAVRASAAIPNV	
	orf137-1 FIKGEKLQNYINRKVGGRQIQQFPIKFAAVATDFETGKAVAFNQGNAGQAVRASAAIPNV	
30	orf137a.pep FQPVIIGRHTYVDGGLSQPVPVSAARRXXXXXVIAVDISARPSKNISQGFFSYLDQTLNV	,
	orf137a.pep MSVSALQNELGQADVVIKPQVLDLGAVGGFDQKKRAIRLGEEAARAALPEIKRKLAAYRY	
35	orf137-1 MSVSALQNELGQADVVIKPQVLDLGAVGGFDQKKRAIRLGEEAARAALPEIKRKLAAYRY	
	Homology with a predicted ORF from N.gonorrhoeae	
	ORF137 shows 89.9% identity over a 149aa overlap with a predicted ORF (ORF137ng	) from
	N.gonorrhoeae:	
40	orf137.pep MENMVTFSKIRPLLAIAAAALLAAXRTAGNNAVRKPVQTAKPAAVVGLALGGGASKGFAH	60
	orf137ng MENMVTFSKIRSFLAIAAAALLAACGTAGNNAARKPVQTAKPAAVVALALGGGASKGFAH	60
45	orf137.pep VGIIKVLKENGIPVKVVTGTSAGSIVGNLFASGMSPDRLELEAEILGKTDLVDLTLSTNG	120.
	orf137ng IGIVKVLKENGIPVKVVTGTSAGSIVGSLLASGMSPDRLELEAEILGKTDLVDLTLSTSG	120
	orf137.pep FIKGAKLONYINRKLRGMQIQOFPIKFAA	149
50	orf137ng FIKGEKLQNYINRKVGGRQIQQFPIKFAAVATDFETGKAVAFNQGNAGQAVRASAAIPNV	180
	The complete length ORF137ng nucleotide sequence <seq 563="" id=""> is:</seq>	
55	1 ATGGAAAATA TGGTAACGTT TTCAAAAATC AGATCATTTT TGGCAATCGC 51 CGCCGCCGCG TTGCTTGCCG CCTGCGGTAC GGCGGGAAAC AATGCCGCCC 101 GCAAGCCGGT GCAAACCGCC AAACCCGCCG CAGTGGTCGC TTTGGCACTC 151 GGTGGCGGCG CATCTAAAGG ATTTGCCCAT ATAGGAATTG TTAAGGTTTT 201 GAAAGAAAAC GGTATTCCTG TGAAGGTGGT TACCGGCACA TCGGCAGGTT 251 CGATAGTCGG CAGCCTTTTG GCATCGGGTA TGTCGCCCGA CCGCCTCGAA 301 TTGGAAGCCG AGATTTTAGG TAAAACCGAT TTAGTCGATT TAACCTTGTC	
60	351 CACCAGTGGT TTTATCAAAG GCGAAAAGCT GCAAAATTAC ATCAACCGAA 401 AAGTCGGCGG CAGGCAGATT CAGCAGTTTC CCATCAAATT TGCCGCCGTT 451 GCCACTGATT TTGAAACCGG CAAGGCCGTC GCTTTCAATC AAGGGAATGC	

	501	CGGGCAGGCG	GTTCGTGCTT	CCGCCGCCAT	TCCCAATGTG	TTCCAGCCAG
	551	TCATCATCGG	CAGGCACAAA	TATGTTGACG	GCGGTCTGTC	GCAGCCCGTG
	601	CCCGTCAGTG	CCGCTCGGCG	GCAGGGGGCG	AATTTCGTGA	TTGCCGTCGA
	651	TATTTCCGCA	CGTCCGAGCA	AAAATGTCGG	TCAAGGTTTC	TTCTCTTATC
5	701	TCGATCAGAC	GCTGAACGTG	ATGAGCGTTT	CCGTGTTGCA	AAACGAGTTG
	751	gggcAGGCGG	ATGTGGTTAT	CAAACCGCag	gtTTTGGATT	TGGGTGCAGT
	801	CGGCGGATTC	GATCAGAAAA	AGCGCGCCAT	CCGGTTGGGC	GAGGAGGCAG
	851	CACGTGCCGC	ATTGCCTGAA	ATCAAACGCA	AACTGGCGGC	ATACCGTTAT
	901	TGA				

This encodes a protein having amino acid sequence <SEQ ID 564>: 10

```
1 MENMVTFSKI RSFLAIAAAA LLAACGTAGN NAARKPVQTA KPAAVVALAL
                51 GGGASKGFAH IGIVKVLKEN GIPVKVVTGT SAGSIVGSLL ASGMSPDRLE
                101
                    LEAEILGKTD LVDLTLSTSG FIKGEKLONY INRKVGGRQI QQFPIKFAAV
                151 ATDFETGKAV AFNQGNAGQA VRASAAIPNV FQPVIIGRHK YVDGGLSQPV
15
                201 PVSAARRQGA NFVIAVDISA RPSKNVGQGF FSYLDQTLNV MSVSVLQNEL
                251
                    GQADVVIKPQ VLDLGAVGGF DQKKRAIRLG EEAARAALPE IKRKLAAYRY
                301
```

### ORF137ng and ORF137-1 show 96.0% identity in 300 aa overlap:

20	orf137ng	MENMVTFSKIRSFLAIAAAALLAACGTAGNNAARKPVQTAKPAAVVALALGGGASKGFAH
	orf137-1	MENMVTFSKIRPLLAIAAAALLAACGTAGNNAVRKPVQTAKPAAVVGLALGGGASKGFAH
	orf137ng	IGIVKVLKENGIPVKVVTGTSAGSIVGSLLASGMSPDRLELEAEILGKTDLVDLTLSTSG:  :
25	orf137-1	VGIIKVLKENGIPVKVVTGTSAGSIVGSLFASGMSPDRLELEAEILGKTDLVDLTLSTSG
	orf137ng	FIKGEKLQNYINRKVGGRQIQQFPIKFAAVATDFETGKAVAFNQGNAGQAVRASAAIPNV
30	orf137-1	FIKGEKLQNYINRKVGGRQIQQFPIKFAAVATDFETGKAVAFNQGNAGQAVRASAAIPNV
50	orf137ng	FQPVIIGRHKYVDGGLSQPVPVSAARRQGANFVIAVDISARPSKNVGQGFFSYLDQTLNV
	orf137-1	FQPVIIGRHTYVDGGLSQPVPVSAARRQGANFVIAVDISARPGKNISQGFFSYLDQTLNV
35	orf137ng	MSVSVLQNELGQADVVIKPQVLDLGAVGGFDQKKRAIRLGEEAARAALPEIKRKLAAYRY
	orf137	MSVSALQNELGQADVVIKPQVLDLGAVGGFDQKKRAIRLGEEAARAALPEIKRKLAAYRY

Based on the presence of a predicted prokaryotic membrane lipoprotein lipid attachment site (underlined) in the gonococcal protein, it is predicted that the proteins from N.meningitidis and N.gonorrhoeae, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 68

40

The following partial DNA sequence was identified in N.meningitidis <SEQ ID 565>:

	1	ATGTTTCGTT	TACAATTCAG	GCTGTTTCCC	CCTTTGCGAA	CCGCCATGCA
45	51	CATCCTGTTG	ACCGCCCTGC	TCAAATGCCT	CTCCCTGcTG	CCGCTTTCCT
	101	GTCTGCACAC	GCTGGGAAAC	CGGCTCGGAC	ATCTGGCGTT	TTACCTTTTA
	151	AAGGAAGACC	GCGCGCGCAT	CGTCGCCmAT	ATGCGGCAGG	CGGGTTTGAA
	201	CCCCGACCCC	AAAACGGTCA	AAGCCGTTTT	TGCGGAAACG	GCAAAAGGCG
	251	GTTTGGAACT	TGCCCCCGCG	TTTTTCAGAA	AACCGGAAGA	CATAGAAACA
50	301	ATGTTCAAAG	CGGTACACGG	CTGGGAACAT	GTGCAGCAGG	CTTTGGACAA
	351	ACACGAAGGG	CTGCTATTC.	•		
50	251 301	CCCCGACCCC GTTTGGAACT ATGTTCAAAG	AAAACGGTCA TGCCCCCGCG CGGTACACGG	AAGCCGTTTT TTTTTCAGAA CTGGGAACAT	TGCGGAAACG AACCGGAAGA	GCAAAAGGCG CATAGAAACA

This corresponds to the amino acid sequence <SEQ ID 566; ORF138>:

- 1 MFRLQFRLFP PLRTAMHILL TALLKCLSLL PLSCLHTLGN RLGHLAFYLL 51 KEDRARIVAX MRQAGLNPDP KTVKAVFAET AKGGLELAPA FFRKPEDIET
- 55 101 MFKAVHGWEH VQQALDKHEG LLF

### Further work revealed the complete nucleotide sequence <SEQ ID 567>:

```
ATGTTTCGTT TACAATTCAG GCTGTTTCCC CCTTTGCGAA CCGCCATGCA
                51
                    CATCCTGTTG ACCGCCCTGC TCAAATGCCT CTCCCTGCTG CCGCTTTCCT
               101
                    GTCTGCACAC GCTGGGAAAC CGGCTCGGAC ATCTGGCGTT TTACCTTTTA
5
                    AAGGAAGACC GCGCGCGCAT CGTCGCCAAT ATGCGGCAGG CGGGTTTGAA
               151
                     CCCCGACCC AAAACGGTCA AAGCCGTTTT TGCGGAAACG GCAAAAGGCG
               201
                     GTTTGGAACT TGCCCCCGCG TTTTTCAGAA AACCGGAAGA CATAGAAACA
               251
                301
                    ATGTTCAAAG CGGTACACGG CTGGGAACAT GTGCAGCAGG CTTTGGACAA
                    ACACGAAGGG CTGCTATTCA TCACGCCGCA CATCGGCAGC TACGATTTGG
                351
10
                401
                     GCGGACGCTA CATCAGCCAG CAGCTTCCGT TCCCGCTGAC CGCCATGTAC
                    AAACCGCCGA AAATCAAAGC GATAGACAAA ATCATGCAGG CGGGCAGGGT
                451
                501
                     TCGCGGCAAA GGAAAAACCG CGCCTACCAG CATACAAGGG GTCAAACAAA
                     TCATCAAAGC CCTGCGTTCG GGCGAAGCAA CCATCGTCCT GCCCGACCAC
                551
                     GTCCCCTCCC CTCAAGAAGG CGGGGAAGGC GTATGGGTGG ATTTCTTCGG
                601
15
                651
                     CAAACCTGCC TATACCATGA CGCTGGCGGC AAAATTGGCA CACGTCAAAG
                701
                     GCGTGAAAAC CCTGTTTTTC TGCTGCGAAC GCCTGCCTGG CGGACAAGGT
                751
                     TTCGATTTGC ACATCCGCCC CGTCCAAGGG GAATTGAACG GCGACAAAGC
                     CCATGATGCC GCCGTGTTCA ACCGCAATGC CGAATATTGG ATACGCCGTT
                801
                851
                     TTCCGACGCA GTATCTGTTT ATGTACAACC GCTACAAAAT GCCGTAA
     This corresponds to the amino acid sequence <SEQ ID 568; ORF138-1>:
20
```

```
1 MFRLQFRLFP PLRTAMHILL TALLKCLSLL PLSCLHTLGN RLGHLAFYLL
51 KEDRARIVAN MRQAGLNPDP KTVKAVFAET AKGGLELAPA FFRKPEDIET
101 MFKAVHGWEH VQQALDKHEG LLFITPHIGS YDLGGRYISQ QLPFPLTAMY
151 KPPKIKAIDK IMQAGRVRGK GKTAPTSIQG VKQIIKALRS GEATIVLPDH
201 VPSPQEGGEG VWVDFFGKPA YTMTLAAKLA HVKGVKTLFF CCERLPGGQG
251 FDLHIRPVQG ELNGDKAHDA AVFNRNAEYW IRRFPTQYLF MYNRYKMP*
```

Computer analysis of this amino acid sequence gave the following results:

### Homology with a predicted ORF from N. meningitidis (strain A)

ORF138 shows 99.2% identity over a 123aa overlap with an ORF (ORF138a) from strain A of N.

#### 30 meningitidis:

25

		. 10	20	30	40	50	60
	orf138.pep	MFRLQFRLFPPLR	TAMHILLTALI	KCLSLLPLS	CLHTLGNRLG	<b>ILAFYLLKE</b> DR	ARIVAX
				411111111			
	orf138a	MFRLQFRLFPPLR	TAMHILLTALI				
35		10	20	30	40	50	60
		70	80	90	100	110	120
	orf138.pep	MRQAGLNPDPKTV					
		11111111111111	11111111111	111111111	14 1 1 1 1 1 1 1 1 1 1 1		1111111
40	orf138a	MRQAGLNPDPKTV	KAVFAETAKGO	ELELAPAFFRI	KPEDIETMFK/	AVHGWEHVQQA	LDKHEG
		70	80	90	100	110	120
	•			•			
	orf138.pep	LLF					
45		111				•	
	orf138a	LLFITPHIGSYDI	GGRYISQQLP	PLTAMYKPPI	KIKAIDKIMQ	AGRVRGKGKT <i>A</i>	APTSIQG
		130	140	150	160	170	180

#### The complete length ORF138a nucleotide sequence <SEQ ID 569> is:

	1	ATGTTTCGTT	TACAATTCAG	GCTGTTTCCC	CCTTTGCGAA	CCGCCATGCA
50	51	CATCCTGTTG	ACCGCCCTGC	TCAAATGCCT	CTCCCTGCTG	CCGCTTTCCT
	101	GTCTGCACAC	GCTGGGAAAC	CGGCTCGGAC	ATCTGGCGTT	TTACCTTTTA
	151	AAGGAAGACC	GCGCGCGCAT	CGTCGCCAAT	ATGCGTCAGG	CAGGCATGAA
*	201	TCCCGACCCC	AAAACGGTCA	AAGCCGTTTT	TGCGGAAACG	GCAAAAGGCG
	251	GTTTGGAACT	TGCCCCCGCG	TTTTTCAGAA	AACCGGAAGA	CATAGAAACA
55	301	ATGTTCAAAG	CGGTACACGG	CTGGGAACAT	GTGCAGCAGG	CTTTGGACAA
	351	ACACGAAGGG	CTGCTATTCA	TCACGCCGCA	CATCGGCAGC	TACGATTTGG
	401	GCGGACGCTA	CATCAGCCAG	CAGCTTCCGT	TCCCGCTGAC	CGCCATGTAC
	451	AAACCGCCGA	AAATCAAAGC	GATAGACAAA	ATCATGCAGG	CGGGCAGGGT
	501	TCGCGGCAAA	GGAAAAACCG	CGCCTACCAG	CATACAAGGG	GTCAAACAAA
60	551	TCATCAAAGC	CCTGCGTTCG	GGCGAAGCAA	CCATCGTCCT	GCCCGACCAC

```
601 GTCCCCTCCC CTCAAGAAGG CGGGGAAGGC GTATGGGTGG ATTTCTTCGG
                  CAAACCTGCC TATACCATGA CGCTGGCGGC AAAATTGGCA CACGTCAAAG
              651
                  GCGTGAAAAC CCTGTTTTTC TGCTGCGAAC GCCTGCCTGG CGGACAAGGT
              701
                  TTCGATTTGC ACATCCGCCC CGTCCAAGGG GAATTGAACG GCGACAAAGC
              751
                  CCATGATGCC GCCGTGTTCA ACCGCAATGC CGAATATTGG ATACGCCGTT
5
              801
              851 TTCCGACGCA GTATCTGTTT ATGTACAACC GCTACAAAAT GCCGTAA
    This encodes a protein having amino acid sequence <SEQ ID 570>:
                1 MFRLQFRLFP PLRTAMHILL TALLKCLSLL PLSCLHTLGN RLGHLAFYLL
                  KEDRARIVAN MRQAGLNPDP KTVKAVFAET AKGGLELAPA FFRKPEDIET
               51
                  MFKAVHGWEH VQQALDKHEG LLFITPHIGS YDLGGRYISQ QLPFPLTAMY
10
              101
                  KPPKIKAIDK IMQAGRVRGK GKTAPTSIQG VKQIIKALRS GEATIVLPDH
              151
                  VPSPQEGGEG VWVDFFGKPA YTMTLAAKLA HVKGVKTLFF CCERLPGGQG
              201
                  FDLHIRPVQG ELNGDKAHDA AVFNRNAEYW IRRFPTQYLF MYNRYKMP*
              251
     ORF138a and ORF138-1 show 99.7% identity over a 298aa overlap:
                      MFRLQFRLFPPLRTAMHILLTALLKCLSLLPLSCLHTLGNRLGHLAFYLLKEDRARIVAN
15
          orf138a.pep
                      MFRLQFRLFPPLRTAMHILLTALLKCLSLLPLSCLHTLGNRLGHLAFYLLKEDRARIVAN
          orf138-1
                      MROAGMNPDPKTVKAVFAETAKGGLELAPAFFRKPEDIETMFKAVHGWEHVQQALDKHEG
          orf138a.pep
                       20
                      MRQAGLNPDPKTVKAVFAETAKGGLELAPAFFRKPEDIETMFKAVHGWEHVQQALDKHEG
          orf138-1
                      LLFITPHIGSYDLGGRYISQQLPFPLTAMYKPPKIKAIDKIMQAGRVRGKGKTAPTSIQG
          orf138a.pep
                       LLFITPHIGSYDLGGRYISOOLPFPLTAMYKPPKIKAIDKIMQAGRVRGKGKTAPTSIQG
25
          orf138-1
                       VKQIIKALRSGEATIVLPDHVPSPQEGGEGVWVDFFGKPAYTMTLAAKLAHVKGVKTLFF
          orf138a.pep
                       VKQIIKALRSGEATIVLPDHVPSPQEGGEGVWVDFFGKPAYTMTLAAKLAHVKGVKTLFF
          orf138-1
30
                       CCERLPGGQGFDLHIRPVQGELNGDKAHDAAVFNRNAEYWIRRFPTQYLFMYNRYKMP
          or:138a.pep
                       CCERLPGGQGFDLHIRPVQGELNGDKAHDAAVFNRNAEYWIRRFPTQYLFMYNRYKMP
          orf138-1
     Homology with a predicted ORF from N. gonorrhoeae
35
     ORF138 shows 94.3% identity over a 123aa overlap with a predicted ORF (ORF138ng) from
     N.gonorrhoeae:
                      MFRLQFRLFPPLRTAMHILLTALLKCLSLLPLSCLHTLGNRLGHLAFYLLKEDRARIVAX
                                                                               60
          orf138.pep
                      MFRLQFRLFPPLRTAMHILLTALLKCLSLLSLSCLHTLGNRLGHLAFYLLKEDRARIVAN
                                                                               60
 40
          orf138ng
                      \mathtt{MRQAGLNPDPKTVKAVFAETAKGGLELAPAFFRKPEDIETMFKAVHGWEHVQQALDKHEG}
          orf138.pep
                      MROAGLNPDTOTVKAVFAETAKCGLELAPAFFKKPEDIETMFKAVHGWEHVQQALDKGEG
                                                                              120
           orf138ng
 45
                                                                              123
           orf138.pep
                      LLF
                      111
                      LLFITPHIGSYDLGGRYISQQLPFHLTAMYKPPKIKAIDKIMQAGRVRGKGKTAPTGIQG 180
           orf138ng
      The complete length ORF138ng nucleotide sequence <SEQ ID 571> is:
 50
                 1 ATGTTTCGTT TACAATTCAG GCTGTTTCCC CCTTTGCGAA CCGCCATGCA
                   CATCCTGTTG ACCGCCCTGC TCAAATGCCT CTCCCTGCTG TCGCTTTCCT
                51
                   GTCTGCACAC GCTGGGAAAC CGGCTCGGAC ATCTGGCGTT TTACCTTTTA
               101
                   AAGGAAGACC GCGCGCGCAT CGTCGCCAAT ATGCGGCAGG CGGGTTTGAA
               151
                   CCCCGACACG CAGACGGTCA AAGCCGTTTT TGCGGAAACG GCAAAATGCG
               201
                   GTTTGGAACT TGCCCCCGCG TTTTTCAAAA AACCGGAAGA CATCGAAACA
 55
               251
                   ATGTTCAAAG CGGTACACGG CTGGGAACAC GTGCAGCAGG CTTTGGACAA
GGGCGAAGGG CTGCTGTTCA TCACGCCGCA CATCGGCAGC TACGATTTGG
               301
               351
                   GCGGACGCTA CATCAGCCAG CAGCTTCCGT TCCACCTGAC CGCCATGTAC
               401
               451 AAGCCGCCGA AAATCAAAGC GATAGACAAA ATCATGCAGG CGGGCAGGGT
                    GCGCGGCAAA GGCAAAACcg cgccaccgg catACAAGGG GTCAAACAAA
 60
               501
               551 tcatcaAGGC CCTGCGCGCG GGCGAGGCAA CCAtcATCCT GCCCGACCAC
```

55

```
GTCCCTTCTC CGCAGGAAGG CGGCGGCGTG TGGGCGGATT TTTTCGGCAA
ACCTGCATAC ACCATGACAC TGGCGGCAAA ATTGGCACAC GTCAAAGGCG
TGAAAACCCT GTTTTTCTGC TGCGAACGCC TGCCCGACGG ACAAGGCTTC
              601
              651
              701
                   GTGTTGCACA TCCGCCCCGT CCAAGGGGAA TTGAACGGCA ACAAAGCCCA
              751
5
              801
                   CGATGCCGCC GTGTTCAACC GCAATACCGA ATATTGGATA CGCCGTTTTC
                   CGACGCAGTA TCTGTTTATG TACAACCGCT ATAAAACGCC GTAA
    This encodes a protein having amino acid sequence <SEQ ID 572>:
                   MFRLQFRLFP PLRTAMHILL TALLKCLSLL SLSCLHTLGN RLGHLAFYLL
                   KEDRARIVAN MRQAGLNPDT QTVKAVFAET AKCGLELAPA FFKKPEDIET
               51
10
              101
                   MFKAVHGWEH VQQALDKGEG LLFITPHIGS YDLGGRYISQ QLPFHLTAMY
                   KPPKIKAIDK IMQAGRVRGK GKTAPTGIQG VKQIIKALRA GEATIILPDH
              151
                   VPSPQEGGGV WADFFGKPAY TMTLAAKLAH VKGVKTLFFC CERLPDGQGF
              201
                  VLHIRPVQGE LNGNKAHDAA VFNRNTEYWI RRFPTQYLFM YNRYKTP*
              251
    ORF138ng and ORF138-1 show 94.3% identity over 299aa overlap:
15
         orf138-1.pep
                       MFRLQFRLFPPLRTAMHILLTALLKCLSLLPLSCLHTLGNRLGHLAFYLLKEDRARIVAN
                       orf138ng
                       MFRLQFRLFPPLRTAMHILLTALLKCLSLLSLSCLHTLGNRLGHLAFYLLKEDRAR IVAN
         orf138-1.pep
                       MRQAGLNPDPKTVKAVFAETAKGGLELAPAFFRKPEDIETMFKAVHGWEHVQQALDKHEG
                       20
                       MRQAGLNPDTQTVKAVFAETAKCGLELAPAFFKKPEDIETMFKAVHGWEHVQQALDKGEG
         orf138ng
         orf138-1.pep
                       {\tt LLFITPHIGSYDLGGRYISQQLPFPLTAMYKPPKIKAIDKIMQAGRVRGKGKTAPTSIQG}
                       25
         orf138ng
                       LLFITPHIGSYDLGGRYISQQLPFHLTAMYKPPKIKAIDKIMQAGRVRGKGKTAPTGIQG
         orf138-1.pep
                       VKQIIKALRSGEATIVLPDHVPSPQEGGEGVWVDFFGKPAYTMTLAAKLAHVKGVKTLFF
                       orf138ng
                       VKQIIKALRAGEATIILPDHVPSPQEGG-GVWADFFGKPAYTMTLAAKLAHVKGVKTLFF
30
          orf138-1.pep
                        CCERLPGGQGFDLHIRPVQGELNGDKAHDAAVFNRNAEYWIRRFPTQYLFMYNRYKMP
                        orf138ng
                        CCERLPDGQGFVLHIRPVQGELNGNKAHDAAVFNRNTEYWIRRFPTQYLFMYNRYKTP
     In addition, ORF138ng is homologous to htrB protein from Pseudomonas fluorescens:
35
          gnl|PID|e334283 (Y14568) htrB [Pseudomonas fluorescens] Length = 253
           Score = 80.8 bits (196), Expect = 9e-15
           Identities = 49/151 (32%), Positives = 79/151 (51%), Gaps = 6/151 (3%)
          Query: 101 MFKAVHGWEHVQQALDKGEGLLFITPHIGSYD-LGGRYISQQLPFHLTAMYKPPKIKAID 159
40
                   + + V G E +++AL G+G++ IT H+G+++ L
                                                     Y SQ P
                                                                  Y+PPK+KA+D
          Sbjct: 94 LVREVEGLEVLKEALASGKGVVGITSHLGNWEVLNHFYCSQCKPI---IFYRPPKLKAVD 150
          Query: 160 KIMQAGRVRGKGKTAPTGIQGVKQIIKALRAGEATIILPDHVPSPQEGGGVWADFFGKPA 219
                    ++++ RV+
                               KA + +G+ +IK+RG
                                                    I D P P E G++ FF
45
          Sbjct: 151 ELLRKQRVQLGNKVAASTKEGILSVIKEVRKGGQVGIPAD--PEPAESAGIFVPFFATQA 208
          Quary: 220 YTMTLAAKLAHVKGVKTLFFCCERLPDGQGF 250
                                         RLPDG G+
          Sbjct: 209 LTSKFVPNMLAGGKAVGVFLHALRLPDGSGY 239
50
     Based on this analysis, including the presence of a putative transmembrane domain in the
```

gonococcal protein, it was predicted that the proteins from N. meningitidis and N. gonorrhoeae, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF138-1 (57kDa) was cloned in the pGex vectors and expressed in E.coli, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 14A shows the results of affinity purification of the GST-fusion protein. Purified GST-fusion protein was used to immunise mice, whose sera were used for ELISA (positive result) and FACS analysis (Figure 14B). These experiments confirm that ORF138-1 is a surface-exposed protein, and that it is a useful immunogen.

#### Example 69

The following partial DNA sequence was identified in N.meningitidis <SEQ ID 573>:

```
5
                   ..GCGTGGTCGG CCGGCGAATC GTGGCGTGTG TTAATGGAAA GTGAAACGTG
               51
                     GCATGCGGTG TGGAATACTT TGCGCTTCTC GGCGGCGGCG GTGTATGCGG
              101
                     151
                     ATGCGCGGGC TGATGTTTTA GCCGTTTATG GTGTCGCCGG TTTGTGTTTC
                     GGCGGGCGTG CTGCTGCTTT ATCCGCAGTG GACGGCTTCG TTGCCGTTGC
              201
10
              251
                     TGCTGGCGAT GTATGCGCTG CTGGCGTATC CGTTTGTGGC AAAAGATGTT
              301
                     TTATCAGCCT GGGATGCACT GCCGCCGGAT TACGGCAGGG CGGCGGGGG
                     TTTGGGTGCA AACGGCTTTC AGACGGCATG CCGCATCACG TTCCCCCTCT
              351
              401
                     TGAAACCGGC GTTGCGGCGC GGTCTGACTT TGGCGGCGGC AACCTGCGTG
              451
                     GGCGAATTTG CGGCGACATT GTTTCTGTCG CGTCCGGAAT GGCAGACGCT
15
               501
                     GACGACTTTG ATTTATGCCT ATTTGGGACG CGCGGGTGAG GATAATTACG
                     CGCGGGCGAT GGTGCTG...
```

This corresponds to the amino acid sequence <SEQ ID 574; ORF139>:

```
1 ...AWSAGESWRV LMESETWHAV WNTLRFSAAA VYAAAVLGVV YAAPARRSAW
51 MRGLMFXPFM VSPVCVSAGV LLLYPQWTAS LPLLLAMYAL LAYPFVAKDV
20 101 LSAWDALPPD YGRAAAGLGA NGFQTACRIT FPLLKPALRR GLTLAAATCV
151 GEFAATLFLS RPEWQTLTTL IYAYLGRAGE DNYARAMVL.
```

Further work revealed the complete nucleotide sequence <SEQ ID 575>:

```
ATGGATGGAC GGCGTTGGGT GGTATGGGGT GCTTTTGCCC TGCTGCCTTC
                     GGCTTTTTTG GCGGTAATGG TCGTTGCGCC TTTGTGGGCG GTGGCGGCGT
25
                    ATGACGGTTT GGCGTGCGC GCGGTGCTGT CGGATGCCTA TATGCTCAAA
                101
                     CGTTTGGCGT GGACGGTATT TCAGGCAGCG GCAACCTGTG TGCTGGTGCT
                    GCCTTTGGGC GTGCCTGTCG CGTGGGTGCT GGCGCGGCTG GCGTTTCCGG
                201
                     GGCGGGCTTT GGTGCTGCGC CTGCTGATGC TGCCTTTTGT GATGCCCACG
                251
                    TTGGTGGCGG GCGTGGGCGT GCTGGCCCTG TTCGGGGCGG ACGGGCTGTT
                301
30
                351
                    GTGGCGCGC AGGCAGGATA CGCCGTATCT GTTGTTGTAC GGCAATGTGT
                    TTTTCAACCT TCCTGTGTTG GTCAGGGCGG CGTATCAGGG GTTTGTGCAA
                401
                     GTGCCTGCGG CACGGCTTCA GACGGCACGG ACGTTGGGCG CGGGGGCGTG
                451
                    GCGGCGGTTT TGGGACATTG AAATGCCCGT TTTGCGCCCG TGGCTTGCCG
                501
                551
                     GCGGCGTGTG CCTTGTCTTT CTGTATTGTT TTTCCGGGTT CGGGCTGGCG
35
                     CTGCTGCTGG GCGGCAGCCG TTATGCCACG GTCGAAGTGG AAATTTACCA
                601
                651
                     GTTGGTCATG TTCGAACTCG ATATGGCGGT TGCTTCGGTG CTGGTGTGGC
                701
                    TGGTGTTGGG GGTAACGGCG GCGGCAGGGT TGCTGTATGC GTGGTTCGGC
                     AGGCGCGCG TTTCGGATAA GGCGGTTTCC CCTGTGATGC CGTCGCCGCC
                751
                801
                    GCAGTCGGTC GGGGAATATG TGCTGCTGGC GTTTGCGGCG GCGGTGTTGT
40
                     CTGTGTGCTG CCTGTTTCCT TTGTTGGCAA TTGTTGTGAA AGCGTGGTCG
                851
                901
                     GCCGGCGAAT CGTGGCGTGT GTTAATGGAA AGTGAAACGT GGCAGGCGGT
                951
                     GTGGAATACT TTGCGCTTCT CGGCGGCGGC GGTGTATGCG GCGGCGGTTT
                     TGGGTGTGGT GTATGCGGCG GCGGCGCGC GGTCGGCGTG GATGCGCGGG
               1001
               1051
                     CTGATGTTTT TGCCGTTTAT GGTGTCGCCG GTTTGTGTTT CGGCGGGCGT
45
               1101
                     GCTGCTGCTT TATCCGCAGT GGACGGCTTC GTTGCCGTTG CTGCTGGCGA
                     TGTATGCGCT GCTGGCGTAT CCGTTTGTGG CAAAAGATGT TTTATCAGCC
               1151
               1201
                     TGGGATGCAC TGCCGCCGGA TTACGGCAGG GCGGCGGCGG GTTTGGGTGC
                     AAACGGCTTT CAGACGGCAT GCCGCATCAC GTTCCCCCTC TTGAAACCGG
               1251
                     CGTTGCGGCG CGGTCTGACT TTGGCGGCGG CAACCTGCGT GGGCGAATTT
               1301
50
                     GCGGCGACAT TGTTTCTGTC GCGTCCGGAA TGGCAGACGC TGACGACTTT
               1351
                     GATTTATGCC TATTTGGGAC GCGCGGGTGA GGATAATTAC GCGCGGGCGA
                     TGGTGCTGAC ATTGCTGTTG GCGGCGTTCG CGCTGGGTAT TTTCCTGCTG
               1451
               1501
                     TTGGACGCC GCGAAGGCGG AAAACAGACG GAAACGTTAT AA
```

This corresponds to the amino acid sequence <SEQ ID 576; ORF139-1>:

55	1	MDGRRWVVWG	AFALLPSAFL	AVMVVAPI.WA	VAAYDGLAWR	AVLSDAYMT.K
<b></b>		RLAWTVFQAA				
	101	LVAGVGVLAL	FGADGLLWRG	RODTPYLLLY	GNVFFNLPVL	VRAAYQGFVQ
	151	VPAARLOTAR	TLGAGAWRRF	WDIEMPVLRP	WLAGGVCLVF	LYCFSGFGLA

1251

1351

```
201 LLLGGSRYAT VEVEIYQLVM FELDMAVASV LVWLVLGVTA AAGLLYAWFG
251 RRAVSDKAVS PVMPSPPQSV GEYVLLAFAA AVLSVCCLFP LLAIVVKAWS
301 AGESWRVLME SETWQAVWNT LRFSAAAVYA AAVLGVVYAA AARRSAWMRG
351 LMFLPFMVSP VCVSAGVLLL YPQWTASLPL LLAMYALLAY PFVAKDVLSA
401 WDALPPDYGR AAAGLGANGF QTACRITFPL LKPALRRGLT LAAATCVGEF
451 AATLFLSRPE WQTLTTLIYA YLGRAGEDNY ARAMVLTLLL AAFALGIFLL
501 LDGGEGGKQT ETL*
```

Computer analysis of this amino acid sequence gave the following results:

### Homology with a predicted ORF from N. meningitidis (strain A)

ORF139 shows 94.7% identity over a 189aa overlap with an ORF (ORF139a) from strain A of N. meningitidis:

	orf139.pep			10 AWSAGESWR	0 20 VLMESETWHAVWNT	30
	• •	•		111111111		111111
15	orf139a	QSVGEYVLLAFAAAVX	SVCCLFXLLAI	<u>VV</u> KAWSAGESWR'	VLMESETWQAVWNT	XRFS <u>AAA</u>
		270 280	290	300	310 320	. —
		40	50	60 7	0 80	90
	orf139.pep	VYAAAVLGVVYAAPAR	RSAWMRGLMFX	PFMVSPVCVSAG	VLLLYPOWTASI.PI	T.T.AMYAT.
20		1111111111111	11111111		<del>[                                    </del>	111111
	orf139a	VYAAAVLGVVYAAAAR	rsawmrglmf <u>l</u>	PFMVSPVCVSAG	<u>VLLL</u> XPQWTAS <u>LPL</u>	LLAMYAL
		330 340	350	360	370 380	,
		100	110	120 130	140	150
25	orf139.pep	LAYPFVAKDVLSAWDA		LGANGFOTACRI	replikpai.rrgi.t	יים ארני עריים מעעניי
				11111111111		111111
	orf139a	LAYPEVAKDVLSAXDA	LPPDYGRAAAG	LGANGFQTACRI'	FPLLKPALRRGLT	LAAATCV
		390 400	410	420	430 440	
30	•	160	170	180 189		
	orf139.pep	GEFAATLFLSRPEWQT				
		11111111 11 1111	1111111111111111	$\Pi^*$ $\Pi$ $\Pi$ $\Pi$ $\Pi$ $\Pi$		
	omf139a	GEFAATLFXSRXEWQT	LTTLIYAYXGR	AGXDNYARAMVL'	<b>TLLLAAFALGXFL</b> L	LDGGEGG
		450 460	470	480	490 500	
35	The complete length	th ORF139a nucleot	ido comuna	CEO ID 67	<b>4.</b> •.	
33	The complete length	ii Old 139a ildeleot	ide sequence		/> 1S:	
	1 A7	GGATGGAC GGCGTTGG	C GGTATGGG	CT CCTTTTTCCC	- MCCMCCCMMA	
	51 G0	SCTTTTTTG GCGGCAAT	GG TCGTTGCG	CC TTTGTGGGC	C TGCTGCCTTC	
	101 A7	rgacggttt ggcgtggc	GC GCGGTGCT	GT CGGATGCCT	TATGCTCAAA	
40	151 CG	STTTGGCGT GGACGGTA	IT TCAGGCAG	CG GCAACCTGT(	TGCTGGTGCT	
40	201 GC	CTTTGGGC GTGCCTGT	CG CGTGGGTG	CT GGCGCGGCT(	GCGTTTCCGG	
	251 GG 301 TT	SCGGGCTTT GGTGCTGC	SC CTGCTGAT	GC TGCCTTTTG	GATGCCCACG	
	351 F1	GGTGGCGG GCGTGGGC GGCGCGGC TGGCAGGA	FA CCCCCTAM	TG TTCGGGGCG(	ACGGCCTGTN	
	401 TT	TTTNACCT TCCTGTGT	IA CUCCUIAN	CT GTTGTTGTA(	GGCAATGTGT	
45	451 GT	GCCTGCGG CACGGCTT	CA GACGGCAC	NG ACATTGGGC	GITTGTGCAA	
	501 GC	CGGCGGTTT TGGGACAT	IG AAATGCCC	GT TTTGCGCCC	TGGCTTGCCG	
	551 GC	GGCGTGTG CCTTGTCT	IC CTGTATTG	TT TTTCGGGGT	CGGGCTGGCA	
	601 TT	GCTGCTGG GCGGCAGC	CG TTATGCCA	CG GTCGAAGTG	AAATTTACCA	•
50	651 GT 701 TG	TTGGTCATG TTCGAACT	CG ATATGGCG	GT TGCTTCGGT	CTNGTGTGGC	
50	701 16 751 AG	GGTGTNGGG GGTAACNG GGCGCGCGG TTTCGGAT	OG GCGGCAGG	GT TGCTGTATG(	GTGGTTCGGC	
	801 GC	CAGTCGGTC GGGGAATA	RG TGCTNCTG	CC CCTGTGATG(	CCCCTCTCT	
	851 CT	GTGTGCTG CCTGTTTC	NT TTGTTGGC	AA TTGTTGTGA	ACCGLIGINGL	
	901 GC	CCGCCGAAT CGTGGCGT(	ST GTTAATGG	AA AGTGAAACGT	GGCAGGCGGT	
55	951 GT	rggaatact ntgcgctt(	CT CGGCGGCG	GC GGTGTATGC	GCGGCGGTTT	
	1001 TO	GGTGTGGT GTATGCGG	ce ececece	GC GGTCGGCGT(	GATGCGCGGG	
	1051 CT 1101 GC	GATGTTTT TGCCGTTT	AT GGTGTCGC	CG GTTTGTGTTT	CGGCGGGCGT	
	1101 GC 1151 TG	CTGCTGCTT NATCCGCA	ST GGACGGCT	rc gttgccgct	CTGCTGGCGA	
60	1201 TG	STATGCGCT GCTGGCGT SNGATGCAC TGCCGCCG	SA TTACCCCA	SG CAAAAGATGT	TTTATCAGCC	
•	1251 AA	ACGGCTTT CAGACGGC	OF TIMEGOLA	3G GCGGCGGCGC	TTGAAACCCC	

AAACGGCTTT CAGACGGCAT GCCGCATCAC GTTCCCCCTC TTGAAACCGG CGTTGCGGCG CGGTCTGACT TTGGCGGCGG CAACCTGCGT GGGCGAATTT

GCGGCAACCT TGTTCNTGTC GCGTCNCGAG TGGCAGACGC TGACGACTTT

WO 99/24578 PCT/IB98/01665

```
GATTTATGCC TATNTGGGAC GCGCGGGTGA NGATAATTAC GCGCGGGCGA
            1451
                 TGGTGCTGAC ATTGCTGTTG GCGGCGTTCG CGCTGGGTAT NTTCCTGCTG
                 TTGGACGGCG GCGAAGGCGG AAAACGGACG GAAACGTTAT AA
            1501
    This encodes a protein having amino acid sequence <SEQ ID 578>:
5
                 MDGRRWAVWG AFALLPSAFL AAMVVAPLWA VAAYDGLAWR AVLSDAYMLK
                 RLAWTVFQAA ATCVLVLPLG VPVAWVLARL AFPGRALVLR LLMLPFVMPT
              51
                 LVAGVGVLAL FGADGLXWRG WQDTPYLLLY GNVFFXLPVL VRAAYQGFVQ
             101
                 VPAARLQTAX TLGAGAWRRF WDIEMPVLRP WLAGGVCLVF LYCFSGFGLA
             151
                 LLLGGSRYAT VEVEIYQLVM FELDMAVASV LVWLVXGVTA AAGLLYAWFG
             201
10
                 RRAVSDKAVS PVMPSPPQSV GEYVLLAFAA AVXSVCCLFX LLAIVVKAWS
             251
                 AGESWRVLME SETWQAVWNT XRFSAAAVYA AAVLGVVYAA AARRSAWMRG
             301
                 LMFLPFMVSP VCVSAGVLLL XPQWTASLPL LLAMYALLAY PFVAKDVLSA XDALPPDYGR AAAGLGANGF QTACRITFPL LKPALRRGLT LAAATCVGEF
             351
             401
                 AATLFXSRXE WQTLTTLIYA YXGRAGXDNY ARAMVLTLLL AAFALGXFLL
             451
15
             501 LDGGEGGKRT ETL*
    ORF139a and ORF139-1 show 96.5% homology over a 514aa overlap:
         orf139a.pep MDGRRWAVWGAFALLPSAFLAAMVVAPLWAVAAYDGLAWRAVLSDAYMLKRLAWTVFQAA
                    MDGRRWVVWGAFALLPSAFLAVMVVAPLWAVAAYDGLAWRAVLSDAYMLKRLAWTVFQAA
         orf139-1
20
                    {\tt ATCVLVLPLGVPVAWVLARLAFPGRALVLRLLMLPFVMPTLVAGVGVLALFGADGLXWRG}
         orf139a.pep
                    orf139-1
                    ATCVLVLPLGVPVAWVLARLAFPGRALVLRLLMLPFVMPTLVAGVGVLALFGADGLLWRG
25
                    WQDTPYLLLYGNVFFXLPVLVRAAYQGFVQVPAARLQTAXTLGAGAWRRFWDIEMPVLRP
         orf139a.pep
                     RQDTPYLLLYGNVFFNLPVLVRAAYQGFVQVPAARLQTARTLGAGAWRRFWDIEMPVLRP
         orf139-1
                    WLAGGVCLVFLYCFSGFGLALLLGGSRYATVEVEIYQLVMFELDMAVASVLVWLVXGVTA
         orf139a.pep
30
                    WLAGGVCLVFLYCFSGFGLALLLGGSRYATVEVEIYQLVMFELDMAVASVLVWLVLGVTA
         orf139-1
                    AAGLLYAWFGRRAVSDKAVSPVMPSPPQSVGEYVLLAFAAAVXSVCCLFXLLAIVVKAWS
         orf139a.pep
                    35
                    AAGLLYAWFGRRAVSDKAVSPVMPSPPQSVGEYVLLAFAAAVLSVCCLFPLLAIVVKAWS
         orf139-1
                    AGESWRVLMESETWQAVWNTXRFSAAAVYAAAVLGVVYAAAARRSAWMRGLMFLPFMVSP
         orf139a.pep
                    AGESWRVLMESETWOAVWNTLRFSAAAVYAAAVLGVVYAAAARRSAWMRGLMFLPFMVSP
         orf139-1
40
                    VCVSAGVLLLXPQWTASLPLLLAMYALLAYPFVAKDVLSAXDALPPDYGRAAAGLGANGF
         orf139a.pep
                    VCVSAGVLLLYPQWTASLPLLLAMYALLAYPFVAKDVLSAWDALPPDYGRAAAGLGANGF
         orf139-1
45
         orf139a.pep
                    QTACRITFPLLKPALRRGLTLAAATCVGEFAATLFXSRXEWQTLTTLIYAYXGRAGXDNY
                    orf139-1
                    QTACRITFPLLKPALRRGLTLAAATCVGEFAATLFLSRPEWQTLTTLIYAYLGRAGEDNY
                    ARAMVLTLLLAAFALGXFLLLDGGEGGKRTETLX
         orf139a.pep
50
                    orf139-1
                    ARAMVLTLLLAAFALGIFLLLDGGEGGKOTETLX
     Homology with a predicted ORF from N.gonorrhoeae
     ORF139 shows 95.2% identity over a 189aa overlap with a predicted ORF (ORF139ng) from
     N.gonorrhoeae:
55
          orf139.pep
                                             AWSAGESWRVLMESETWHAVWNTLRFSAAA
                                                                         30
                                             orf139ng
                    QSVGEYVLLAFSVAVLSVCCLFPLSAIVVKAWSAGESRRVLMESETWQAVWNTLRFSAAA
                                                                        327
 60
          orf139.pep
                    VYAAAVLGVVYAAPARRSAWMRGLMFXPFMVSPVCVSAGVLLLYPQWTASLPLLLAMYAL
                                                                         90
```

VFAAAVLGVVYAAAARRLVWMRGLVFLPFMVSPVCVSAGVLLLYPGWTASLPLLLAMYAL

387

orfl39ng

189

```
orf139.pep
                          LAYPFVAKDVLSAWDALPPDYGRAAAGLGANGFQTACRITFPLLKPALRRGLTLAAATCV
                          LAYPFVAKDVLSAWDALPPDYGRAAAGLGANGFQTACRITFPLLKPALRRGLTLAAATCV
           orf139ng
           orf139.pep
                          GEFAATLFLSRPEWQTLTTLIYAYLGRAGEDNYARAMVL
                          orf139ng
                          GEFAATLFLSRPEWQTLTTLIYAYLGRAGEDNYARAMVLTLLLSAFAVCIFLLLDNGEGG
      The complete length ORF139ng nucleotide sequence <SEQ ID 579> is predicted to encode a
      protein having amino acid sequence <SEO ID 580>:
10
                      MDGRCWAVRG AFSLLPSAFL AVMVVAPLWA VAAYDGLAWR AVLSDAYMLK
                      RLAWTVFQAA ATCVLVLPLG VPVAWVLARL AFPGRALVLR LLMLPFVMPT
                 101
                      LVAGVGVLAL FGADGLLWRG RQDTPYLLLY GNVFFNLPVL VRAAYQGFAQ
                      VPAARLQTAR TLGAGAWRPF WDIEMPVLRP WLAGGVCLVF LYCFSGFGLA
LLLGGSRYAT VEVEIYQLVM FELDMAGASA LVWLVLGVTA AAGLLYAWFG
                 151
                 201
15
                      RRAVSDKAVS PVMPSPPQSV GEYVLLAFSV AVLSVCCLFP LSAIVVKAWS
                 251
                 301
                      AGESRRVLME SETWQAVWNT LRFSAAAVFA AAVLGVVYAA AARRLVWMRG
                      LVFLPFMVSP VCVSAGVLLL YPGWTASLPL LLAMYALLAY PFVAKDVLSA
                 351
                      WDALPPDYGR AAAGLGANGF QTACRITFPL LKPALRRGLT LAAATCVGEF
                 401
                       AATLFLSRPE WOTLTTLIYA YLGRAGEDNY ARAMVLTLLL SAFAVCIFLL
                 451
20
                       LDNGEGGKRT ETL*
      Further work revealed a variant gonococcal DNA sequence <SEQ ID 581>:
                      ATGGATGGAC GGTGTTGGGC GGTACGGGGT GCTTTTTCCC TGCTGCCTTC
                  51
                       GGCTTTTTTG GCGGTAATGG TCGTTGCGCC TTTGTGGGCG GTGGCGGCGT
                      ATGACGGTTT GGCGTGCGC GCGGTGCTGT CGGATGCCTA TATGCTCAAA CGTTTGGCGT GGACGGTGTT TCAGGCGGCG GCAACCTGTG TGCTGGTGCT
                 101
25
                 151
                 201
                       GCCTTTGGGC GTGCCTGTCG CGTGGGTGCT GGCGCGGCTG GCGTTCCCGG
                 251
                       GGCGGGCTTT GGTGCTGCGC CTGCTGATGC TGCCGTTTGT GATGCCCACG
                 301
                       CTGGTGGCGG GCGTGGGCGT GCTGGCTCTG TTCGGGGCGG ACGGGCTGTT
                      GTGGCGCGC CGGCAGGATA CGCCGTATCT GTTGTTGTAC GGCAATGTGT
                 351
30
                      TTTTCAACCT GCCCGTGTTG GTCAGGGCGG CGTATCAGGG GTTTGCTCAA
                 401
                 451
                       GTGCCTGCGG CACGGCTTCA GACGGCACGG ACGTTGGGCG CGGGGGCGTG
                      GCGGCGTTT TGGGACATTG AAATGCCCGT TTTGCGCCCG TGGCTTGCCG
GCGGCGTGTG CCTTGTCTTC CTGTATTGTT TTTCGGGGTT CGGGCTGGCA
                 501
                 551
                 601
                      TTGCTGTTGG GCGGCAGCCG TTATGCCACG GTCGAAGTGG AAATTTACCA
35
                 651
                      GTTGGTTATG TTCGAACTCG ATATGGCGGG GGCTTCGGCG CTGGTGTGGC
                      TGGTGTTGGG GGTAACGGCG GCGGCAGGGT TGCTGTATGC GTGGTTCGGC AGGCGCGCG TTTCGGATAA GGCGGTTTCC CCCGTGATGC CGTCGCCGCC
                 701
                 751
                 801
                      GCAATCGGTG GGGGAATATG TATTGCTGGC ATTTTCGGTG GCGGTGTTGT
                 851
                      CCGTGTGCTG CCTGTTTCCT TTGTCGGCAA TTGTTGTGAA AGCGTGGTCG
40
                 901
                      GCCGGCGAAT CGCGGCGTGT GTTAATGGAA AGTGAAACGT GGCAGGCAGT
                      GTGGAATACt ttGCGCTTTT CGGCGGCGGC GGTGTTTGCG GCGGCGGTTT
                 951
                1001
                       TGGGTGTGGT GTATGCGGCG GCGGCGCGGC GGCTGGTGTG GATGCGCGGA
                1051
                       CTGGTGTTTT TACCGTTTAT GGTGTCGCCG GTTTGTGTTT CGGCGGGCGT
                1101
                       GCTGCTGCTT TATCCGGGGT GGACGGCTTC GTTACCGCTG CTGCTGGCGA
45
                       TGTATGCGCT GCTGGCGTAT CCGTTTGTGG CAAAAGATGT TTTATCGGCC
                1151
                1201
                       TGGGATGCAC TGCCGCCGGA TTACGGCAGG GCGGCGGCAG GTTTGGGCGC
                      AAACGGCTTT CAGACGGCAT GCCGTATCAC GTTCCCCCTC TTGAAACCGG
                1251
                       CGTTGCGGCG CGGTCTGACT TTGGCGGCGG CGACGTGTGT GGGCGAATTT
                1301
                       GCGGCAACCT TGTTCCTGTC GCGTCCGGAA TGGCAGACGT TGACGACTTT
                1351
50
                       GATTTATGCC TATTTGGGGC GTGCGGGTGA GGACAATTAT GCGCGGGCAA
                1401
                1451
                       TGGTGTTGAC ATTGCTGTTG TCGGCATTTG CGGTGTGCAT TTTCCTGCTG
                      TTGGACAACG GCGAAGGCGg aaaACGGACG GAAACGTTAT AA
                1501
      This corresponds to the amino acid sequence <SEQ ID 582; ORF139ng-1>:
                      MDGRCWAVRG AFSLLPSAFL AVMVVAPLWA VAAYDGLAWR AVLSDAYMLK
55
                      RLAWTVFQAA ATCVLVLPLG VPVAWVLARL AFPGRALVLR LLMLPFVMPT
                      LVAGVGVLAL FGADGLLWRG RQDTPYLLLY GNVFFNLPVL VRAAYQGFAQ
VPAARLQTAR TLGAGAWRRF WDIEMPVLRP WLAGGVCLVF LYCFSGFGLA
                 101
                       LLLGGSRYAT VEVEIYQLVM FELDMAGASA LVWLVLGVTA AAGLLYAWFG
                 201
                      RRAVSDKAVS PVMPSPPQSV GEYVLLAFSV AVLSVCCLFP LSAIVVKAWS
                 251
                      AGESRRVLME SETWQAVWNT LRFSAAAVFA AAVLGVVYAA AARRLVWMRG
LVFLPFMVSP VCVSAGVLLL YPGWTASLPL LLAMYALLAY PFVAKDVLSA
WDALPPDYGR AAAGLGANGF QTACRITFPL LKPALRRGLT LAAATCVGEF
60
                 301
                 351
                 401
                 451
                      AATLFLSRPE WQTLTTLIYA YLGRAGEDNY ARAMVLTLLL SAFAVCIFLL
```

LDNGEGGKRT ETL\*

### ORF139ng-1 and ORF139-1 show 95.9% identity over 513aa overlap:

	orf139ng	MDGRCWAVRGAFSLLPSAFLAVMVVAPLWAVAAYDGLAWRAVLSDAYMLKRLAWTVFQAA
5	orf139-1	MDGRRWVVWGAFALLPSAFLAVMVVAPLWAVAAYDGLAWRAVLSDAYMLKRLAWTVFQAA
J	orf139ng	ATCVLVLPLGVPVAWVLARLAFPGRALVLRLLMLPFVMPTLVAGVGVLALFGADGLLWRG
	orf139-1	ATCVLVLPLGVPVAWVLARLAFPGRALVLRLLMLPFVMPTLVAGVGVLALFGADGLLWRG
10	orf139ng	RQDTPYLLLYGNVFFNLPVLVRAAYQGFAQVPAARLQTARTLGAGAWRRFWDIEMPVLRP
	orf139-1	RQDTPYLLLYGNVFFNLPVLVRAAYQGFVQVPAARLQTARTLGAGAWRRFWDIEMPVLRP
15	orf139ng	WLAGGVCLVFLYCFSGFGLALLLGGSRYATVEVEIYQLVMFELDMAGASALVWLVLGVTA
13	orf139-1	WLAGGVCLVFLYCFSGFGLALLLGGSRYATVEVEIYQLVMFELDMAVASVLVWLVLGVTA
	orf139ng	AAGLLYAWFGRRAVSDKAVSPVMPSPPQSVGEYVLLAFSVAVLSVCCLFPLSAIVVKAWS
20	orf139-1	AAGLLYAWFGRRAVSDKAVSPVMPSPPQSVGEYVLLAFAAAVLSVCCLFPLLAIVVKAWS
	orf139ng	AGESRRVLMESETWOAVWNTLRFSAAAVFAAAVLGVVYAAAARRLVWMRGLVFLPFMVSP
25	orf139	AGESWRVLMESETWQAVWNTLRFSAAAVYAAAVLGVVYAAAARRSAWMRGLMFLPFMVSP
	orf139ng	VCVSAGVLLLYPGWTASLPLLLAMYALLAYPFVAKDVLSAWDALPPDYGRAAAGLGANGF
	orf139-1	VCVSAGVLLLYPQWTASLPLLLAMYALLAYPFVAKDVLSAWDALPPDYGRAAAGLGANGF
30	orf139ng	QTACRITFPLLKPALRRGLTLAAATCVGEFAATLFLSRPEWQTLTTLIYAYLGRAGEDNY
	orf139-1	QTACRITFPLLKPALRRGLTLAAATCVGEFAATLFLSRPEWQTLTTLIYAYLGRAGEDNY
35	orf139ng	ARAMVLTLLLSAFAVCIFLLLDNGEGGKRTETL
33	orf139-1	ARAMVLTLLLAAFALGIFLLLDGGEGGKQTETL

Based on the presence of a predicted binding-protein-dependent transport systems inner membrane component signature (underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 70

40

The following partial DNA sequence was identified in N.meningitidis <SEQ ID 583>:

	1	ATGGACGGCT	GGACACAGAC	GCTGTCCGCG	CAAACCCTGT	TGGGCATTTC
	51	GGCGGCGGCA	ATCATCCTCA	TTCTGATTTT	AATCGTCAGA	TTCCGCATCC
45	101	ACGCGCTGCT	GACACTGGTC	ATCGTCAGCC	TGCTGACGGC	TTTGGCAACC
	151	GGTTTGCCCA	CAGGCAGCAT	TGTCAAAGAC	ATACTGGTCA	AAAACTTCGG
	201	CGGCACGCTC	GGCGGCGTGG	CGCTTCTGGT	CGGCCTGGGC	GCGATGCTCG
	251	AACGTTTGGT	C			

This corresponds to the amino acid sequence <SEQ ID 584; ORF140>:

1 MDGWTQTLSA QTLLGISAAA IILILILIVR FRIHALLTLV IVSLLTALAT 51 GLPTGSIVKD ILVKNFGGTL GGVALLVGLG AMLERLV..

Further work revealed the complete nucleotide sequence <SEQ ID 585>:

	1	ATGGACGGCT	GGACACAGAC	GCTGTCCGCG	CAAACCCTGT	TGGGCATTTC
	51	GGCGGCGGCA	ATCATCCTCA	TTCTGATTTT	AATCGTCAAA	TTCCGCATCC
55	101	ACGCGCTGCT	GACACTGGTC	ATCGTCAGCC	TGCTGACGGC	TTTGGCAACC
	151	GGTTTGCCCA	CAGGCAGCAT	TGTCAACGAC	ATACTGGTCA	AAAACTTCGG

```
CGGCACGCTC GGCGGCGTGG CGCTTCTGGT CGGCCTGGGC GCGATGCTCG
                      GACGTTTGGT CGAAACATCC GGCGGCGCAC AGTCGCTGGC GGACGCGCTG
                      ATCCGGATGT TCGGCGAAAA ACGCGCACCG TTCGCGCTGG GCGTTGCCTC
                 301
                      GCTGATTTTC GGCTTCCCGA TTTTCTTCGA TGCCGGACTA ATCGTCATGC
                 351
 5
                      TGCCCATCGT GTTCGCCACC GCACGGCGCA TGAAACAGGA CGTACTGCCC
                 401
                 451
                      TTCGCGCTTG CCTCCATCGG CGCATTTTCC GTCATGCACG TCTTCCTGCC
                 501
                      GCCCCATCCG GGCCCGATTG CCGCTTCCGA ATTTTACGGC GCGAACATCG
                      GCCAAGTTTT GATTTTGGGT CTGCCGACCG CCTTCATCAC ATGGTATTTC
                 551
                 601
                      AGCGGCTATA TGCTCGGCAA AGTGTTGGGG CGCACCATCC ATGTTCCCGT
10
                 651
                      TCCCGAACTG CTCAGCGGCG GCACGCAAGA CAACGACCTG CCGAAAGAAC
                 701
                      CTGCCAAAGC AGGAACGGTC GTCGCCATCA TGCTGATTCC CATGCTGCTG
                      ATTTTCCTGA ATACCGGCGT ATCGGCCCTC ATCAGCGAAA AACTCGTAAG
                 751
                 801
                      TGCGGACGAA ACCTGGGTTC AGACGGCAAA AATAATCGGT TCGACACCGA
                 851
                      TCGCCCTTCT GATTTCCGTA TTGGTCGCAC TGTTTGTCTT GGGACGCAAA
15
                 901
                      CGCGGCGAAA GCGGCAGCGC GTTGGAAAAA ACCGTGGACG GCGCACTCGC
                 951
                      CCCCGTCTGT TCCGTGATTC TGATTACCGG CGCGGGCGGT ATGTTCGGCG
                      GCGTTTTGCG CGCTTCCGGC ATCGGCAAGG CACTCGCCGA CAGCATGGCG
                1001
                      GATTTGGGCA TTCCCGTCCT TTTGGGCTGT TTCCTTGTCG CCTTGGCACT
               1051
                1101
                      GCGTATCGCG CAAGGTTCGG CAACCGTCGC CCTGACCACC GCCGCCGCGC
20
                      TGATGGCTCC TGCCGTTGCC GCCGCCGGCT TTACCGACTG GCAGCTCGCC
                1151
                1201
                      TGTATCGTAT TGGCAACGGC GGCAGGTTCG GTCGGTTGCA GCCACTTCAA
                      CGACTCCGGC TTCTGGCTGG TCGGCCGTCT CTTGGACATG GACGTACCGA
               1251
                1301
                      CCACGCTGAA AACCTGGACG GTCAACCAAA CCCTCATCGC ACTCATCGGC
                      TTTGCCTTGT CCGCACTGCT GTTCGCCATC GTCTGA
                1351
25
      This corresponds to the amino acid sequence <SEQ ID 586; ORF140-1>:
                      MDGWTQTLSA QTLLGISAAA IILILILIVK FRIHALLTLV IVSLLTALAT
                      GLPTGSIVND ILVKNFGGTL GGVALLVGLG AMLGRLVETS GGAQSLADAL IRMFGEKRAP FALGVASLIF GFPIFFDAGL IVMLPIVFAT ARRMKQDVLP FALASIGAFS VMHVFLPPHP GPIAASEFYG ANIGQVLILG LPTAFITWYF
                  51
                 101
                 151
30
                 201
                      SGYMLGKVLG RTIHVPVPEL LSGGTQDNDL PKEPAKAGTV VAIMLIPMLL
                 251
                      IFLNTGVSAL ISEKLVSADE TWVQTAKIIG STPIALLISV LVALFVLGRK
                 301
                      RGESGSALEK TVDGALAPVC SVILITGAGG MFGGVLRASG IGKALADSMA
                 351
                      DLGIPVLLGC FLVALALRIA QGSATVALTT AAALMAPAVA AAGFTDWQLA
                      CIVLATAAGS VGCSHFNDSG FWLVGRLLDM DVPTTLKTWT VNQTLIALIG
                 401
35
                 451
                      FALSALLFAI V*
```

Computer analysis of this amino acid sequence gave the following results:

### Homology with a predicted ORF from N. meningitidis (strain A)

ORF140 shows 95.4% identity over a 87aa overlap with an ORF (ORF140a) from strain A of N. meningicidis:

```
40
                        10
                                        30
                                                40
        orf140.pep
                  MDGWTQTLSAQTLLGISAAAIILILILIVRFRIHALLTLVIVSLLTALATGLPTGSIVKD
                  orf140a
                  MDGWTQTLSAQTLLGISAAAIILILILIVKFRIHALLTLVIVSLLTALATGLPTGSIVND
                                                        50
45
                        70
                                80
        orf140.pep
                  ILVKNFGGTLGGVALLVGLGAMLERLV
                  orf140a
                  VLVKNFGGTLGGVALLVGLGAMLGRLVETSGGAQSLADALIRMFGEKRAPFALGVASLIF
50
                        70
                                80
                                        90
```

The complete length ORF140a nucleotide sequence <SEQ ID 587> is:

	1	ATGGACGGCT	GGACACAGAC	GCTGTCCGCG	CAAACCCTGT	TGGGCATTTC
	51	GGCGGCGGCA	ATCATCCTCA	TTCTGATTTT	AATCGTCAAA	TTCCGCATCC
	101	ACGCGCTGCT	GACACTGGTC	ATCGTCAGCC	TGCTGACGGC	TTTGGCAACC
55	151	GGTTTGCCCA	CAGGCAGCAT	TGTCAACGAC	GTACTGGTCA	AAAACTTCGG
	201	CGGCACGCTC	GCCGCCTGG	CGCTTCTGGT	CGGCCTGGGC	GCGATGCTCG
	251	GACGTTTGGT	CGAAACATCC	GGCGGCGCAC	AGTCGCTGGC	GGACGCGCTG
	301	ATCCGGATGT	TCGGCGAAAA	ACGCGCACCG	TTCGCGCTGG	GCGTTGCCTC
	351			TTTTCTTCGA		
60	401			GCACGGCGCA		
	451			CGCATTTTCC		

	501 GC	CCCCATCCG	GGCCCGATTG	CCGCTTCCGA	ATTTTACGGC	GCGAACATCG	
					CCTTCATCAC		
					CGCACCATCC		
_					CAACGACCTG		
5					TGCTGATTCC		
					ATCAGCGAAA		
					AATAATCGGT		•
					TGTTTGTCTT		
10					ACCGTGGACG CGCGGGCGGT		
10					CACTCGCCGA		
					TTCCTTGTCG		
					CCTGACCACC		
					TTACCGACTG		
15					GTCGGTTGCA		
•					CTTGGACATG		
					CCCTCATCGC		
				GTTCGCCATC			
	This encodes a pro	tein havin	g amino acid	l sequence <	SEO ID 588	· >:	
20	-			-	•		
20					FRIHALLTLV AMLGRLVETS		
					IVMLPIVEAT		
					ANIGQVLILG		
					PKEPAKAGTV		
25					STPIALLISV		
					MFGGVLRASG		
	351 DI	LGIPVLLGC	FLVALALRIA	QGSATVALTT	AAALMAPAVA	AAGFTDWQLA	
	401 C	IVLATAAGS	VGCSHFNDSG	FWLVGRLLDM	DVPTTLKTWT	VNQTLIALIG	
	451 <u>F</u> Z	ALSALLFAI	<u>v</u> *			<del></del>	
		C1 40 1 -L-	00 00/ :4	4:4	4611		
30	ORF140a and ORI	r 140-1 snc	)W 99.8% 10	ennty over a	401aa overi	ap:	
30				•		•	rcetuain 60
30	ORF140a and ORF	p MDGWT(	OTLSAQTLLGI:	SAAAIILILIL	IVKFRIHALLT:	- LVIVSLLTALATGLP1	
30	orf140-1.pep	P MDGWT(	QTLSAQTLLGI 	SAAAIILILIL 	IVKFRIHALLT:	- LVIVSLLTALATGLP1 	111111
30		P MDGWT(	QTLSAQTLLGI 	SAAAIILILIL 	IVKFRIHALLT:	- LVIVSLLTALATGLP1	111111
	orf140-1.pep	P MDGWT(	QTLSAQTLLGI            TLSAQTLLGI	SAAAIILILIL 	IVKFRIHALLT            VKFRIHALLT	LVIVSLLTALATGLP1 	 rgsivnd 60
30 35	orf140-1.pep	p MDGWT(       MDGWT( p ILVKN)	OTLSAOTLLGI                         OTLSAOTLLGI   FGGTLGGVALL 	SAAAIILILIL 	IVKFRIHALLT            IVKFRIHALLT  ETSGGAQSLAD	LVIVSLLTALATGLP1 	GVASLIF 120
	orf140-1.pep	p MDGWT(       MDGWT( p ILVKN)	OTLSAOTLLGI                         OTLSAOTLLGI   FGGTLGGVALL 	SAAAIILILIL 	IVKFRIHALLT            IVKFRIHALLT  ETSGGAQSLAD	LVIVSLLTALATGLP7                 VIVSLLTALATGLP7 ALIRMFGEKRAPFALC	GVASLIF 120
	orf140-1.pep orf140a orf140-1.pep orf140a	p MDGWT(               MDGWT(   MDGWT(   I LVKN)   :           VLVKN)	OTLSAOTLLGI                       OTLSAOTLLGI   FGGTLGGVALL                         FGGTLGGVALL	SAAAIILILIL 	IVKFRIHALLT:             IVKFRIHALLT: ETSGGAQSLAD:           ETSGGAQSLAD:	LVIVSLLTALATGLPT	GUILLI I I I I I I I I I I I I I I I I I
35	orf140-1.pep orf140a orf140-1.pep	P MDGWT(               MDGWT(   MDGWT(	OTLSAOTLLGI                       OTLSAOTLLGI   FGGTLGGVALL                       FGGTLGGVALL   FDAGLIVMLPI	SAAAIILILIL	IVKFRIHALLT:             VKFRIHALLT: ETSGGAQSLAD:           ETSGGAQSLAD: VLPFALASIGA	LVIVSLLTALATGLPT	GUND 60  GVASLIF 120  HILLIH  GVASLIF 120  AASEFYG 180
	orf140-1.pep orf140a orf140-1.pep orf140a orf140-1.pep	p MDGWT(                 MDGWT(   MDGWT(	QTLSAQTLLGI             TLSAQTLLGI             FGGTLGGVALL  FDAGLIVMLPI 	SAAAIILILIL	IVKFRIHALLT:            IVKFRIHALLT: ETSGGAQSLAD:           ETSGGAQSLAD: VLPFALASIGA	LVIVSLLTALATGLPT	FUND 60  EVASLIF 120  HILLIH  EVASLIF 120  AASEFYG 180
35	orf140-1.pep orf140a orf140-1.pep orf140a	p MDGWT(                 MDGWT(   MDGWT(	QTLSAQTLLGI             TLSAQTLLGI             FGGTLGGVALL  FDAGLIVMLPI 	SAAAIILILIL	IVKFRIHALLT:            IVKFRIHALLT: ETSGGAQSLAD:           ETSGGAQSLAD: VLPFALASIGA	LVIVSLLTALATGLPT	FUND 60  EVASLIF 120  HILLIH  EVASLIF 120  AASEFYG 180
35	orf140-1.pep orf140a orf140-1.pep orf140a orf140-1.pep	p MDGWT(                 MDGWT(   MDGWT(	OTLSAQTLLGI                       	SAAAIILILIL	IVKFRIHALLT:	LVIVSLLTALATGLPT	GYASLIF 120 HILLIH FYASLIF 120 AASEFYG 180 HILLIH AASEFYG 810
35	orf140-1.pep orf140a orf140-1.pep orf140a orf140-1.pep	p MDGWT(               MDGWT(    MDGWT(    MDGWT(    MDGWT(    MDGWT(    MDGWT(    MDGWT(    MDGWT(    MDGWT(    MDGWT(     MDGWT(     MDGWT(     MDGWT(      MDGWT(      MDGWT(       MDGWT(	QTLSAQTLLGI             TLSAQTLLGI             FGGTLGGVALL           FDAGLIVMLPI            FDAGLIVMLPI	SAAAIILILIL	IVKFRIHALLT:	LVIVSLLTALATGLPT	FYASLIF 120 ASSEFYG 180 AASEFYG 810 PAKAGTV 240
35 40	orf140-1.pep orf140a orf140-1.pep orf140a orf140-1.pep orf140a	p MDGWT(                 MDGWT(   MDGWT(   P             VLVKN)   GFPIF!   GFPIF!   GFPIF!   P ANIGQ'	QTLSAQTLLGI                       	SAAAIILILIL	IVKFRIHALLT:	LVIVSLLTALATGLPT	PAKAGTV 240
35	orf140-1.pep orf140a orf140-1.pep orf140a orf140-1.pep	p MDGWT(                 MDGWT(   MDGWT(   P             VLVKN)   GFPIF!   GFPIF!   GFPIF!   P ANIGQ'	QTLSAQTLLGI                       	SAAAIILILIL	IVKFRIHALLT:	LVIVSLLTALATGLPT	PAKAGTV 240
35 40	orf140-1.pep orf140a orf140-1.pep orf140a orf140-1.pep orf140a	p MDGWT(                 MDGWT(    MDGWT(   MDGWT(    MDGWT(    MDGWT(    MDGWT(    MDGWT(    MDGWT(    MDGWT(    MDGWT(    MDGWT(    MDGWT(    MDGWT(    MDGWT(    MDGWT(    MDGWT(     MDGWT(     MDGWT(     MDGWT(      MDGWT(      MDGWT(	QTLSAQTLLGI                       	SAAAIILILIL	IVKFRIHALLT:	LVIVSLLTALATGLPT	PAKAGTV 240
35 40	orf140-1.pep orf140a orf140-1.pep orf140a orf140a orf140-1.pep orf140a	p MDGWT(               MDGWT(	QTLSAQTLLGI:              TLSAQTLLGI:                FGGTLGGVALL:                FDAGLIVMLPI:              VLILGLPTAFI:              VLILGLPTAFI:	SAAAIILILIL	IVKFRIHALLT:	LVIVSLLTALATGLPT	PAKAGTV 240 LFVLGRK 300
35 40 45	orf140-1.pep orf140a orf140-1.pep orf140a orf140a orf140-1.pep orf140a	p MDGWT(                 MDGWT(               P   ILVKNI   VLVKNI  P   GFPIFI   GFPIFI   GFPIFI   ANIGQ   ANIGQ	QTLSAQTLLGI	SAAAIILILIL               SAAAIILILIL VGLGAMLGRLV             VGLGAMLGRLV VFATARRMKQD              VFATARRMKQD TWYFSGYMLGK            TWYFSGYMLGK VSALISEKLVS	IVKFRIHALLT:	LVIVSLLTALATGLPT	PAKAGTV 240 LFVLGRK 300
35 40	orf140-1.pep orf140a orf140-1.pep orf140a orf140-1.pep orf140a orf140-1.pep orf140a	p MDGWT(               MDGWT(               MDGWT(             P   ILVKNI   VLVKNI   P   GFPIFI   GFPIFI   GFPIFI   ANIGQ   I           ANIGQ	QTLSAQTLLGI	SAAAIILILIL	IVKFRIHALLT:	LVIVSLLTALATGLPT	GSIVND 60  GVASLIF 120  IIIIIIII  GVASLIF 120  AASEFYG 180  IIIIIII  AASEFYG 810  PAKAGTV 240  IIIIIII  PAKAGTV 240  LFVLGRK 300  IIIIIII  LFVLGRK 300
35 40 45	orf140-1.per orf140a orf140-1.per orf140a orf140-1.per orf140a orf140a orf140a	p MDGWT(               MDGWT(               MDGWT(             P ILVKNI   VLVKNI   VLVKNI   P GFPIFI   GFPIFI   GFPIFI   ANIGO   IIIII     ANIGO   VAIML   VAIML   P RGESG	QTLSAQTLLGI                       	SAAAIILILIL	IVKFRIHALLT:	LVIVSLLTALATGLPT	PAKAGTV 240 LFVLGRK 300 LFVLGRK 300 LFVLGRK 360 LFVLGRK 360 LFVLGRK 360 LFVLGRK 360
35 40 45	orf140-1.pep orf140a orf140-1.pep orf140a orf140-1.pep orf140a orf140-1.pep orf140a orf140-1.pep	P MDGWT(                 MDGWT(               MDGWT(               P ILVKNI   VLVKNI   P GFPIF!   GFPIF!   GFPIF!   ANIGO   I           ANIGO   VAIML   VAIML   P RGESG	QTLSAQTLLGI                       	SAAAIILILIL	IVKFRIHALLT:	LVIVSLLTALATGLPT	PAKAGTV 240 LEVLGRK 300 LEVLLGC 360 LIVILIG 360 LIVILGC 360 LIVILGC 360
35 40 45	orf140-1.pep orf140a orf140-1.pep orf140a orf140-1.pep orf140a orf140-1.pep orf140a	P MDGWT(                 MDGWT(               MDGWT(               P ILVKNI   VLVKNI   P GFPIF!   GFPIF!   GFPIF!   ANIGO   I           ANIGO   VAIML   VAIML   P RGESG	QTLSAQTLLGI                       	SAAAIILILIL	IVKFRIHALLT:	LVIVSLLTALATGLPT	PAKAGTV 240 LEVLGRK 300 LEVLLGC 360 LIVILIG 360 LIVILGC 360 LIVILGC 360
35 40 45 50	orf140-1.per orf140a orf140-1.per orf140a orf140-1.per orf140a orf140-1.per orf140a orf140-1.per orf140a	P MDGWT(                 MDGWT(               MDGWT(               P ILVKNI   VLVKNI   GFPIF!   GFPIF!   ANIGQ'               ANIGQ'   VAIML!   VAIML!   P RGESG:               RGESG:	QTLSAQTLLGI.	SAAAIILILIL	IVKFRIHALLT:	LVIVSLLTALATGLPT	PAKAGTV 240 LFVLGRK 300 LFVLGRK 300 LIVLLGC 360 LIVLLGC 360 LIVLLGC 360
35 40 45	orf140-1.pep orf140a orf140-1.pep orf140a orf140-1.pep orf140a orf140-1.pep orf140a orf140-1.pep	P MDGWT(                 MDGWT(               MDGWT(               P ILVKNI   GFPIF!   GFPIF!   P ANIGQ'               ANIGQ'   VAIML:   VAIML:   P RGESG:   I           RGESG:	QTLSAQTLLGI.	SAAAIILILIL	IVKFRIHALLT:	LVIVSLLTALATGLPT	PAKAGTV 240 LIFVLGRK 300 LIFVLGC 360 EVASLIF 120 AASEFYG 180 LIFVLGRK 300 LIFVLGC 360 LIFVLGC 360 SHFNDSG 420
35 40 45 50	orf140-1.per orf140a orf140-1.per orf140a orf140-1.per orf140a orf140-1.per orf140a orf140-1.per orf140a orf140-1.per	P MDGWT(                 MDGWT(                 MDGWT(               P ILVKNI   GFPIF!   GFPIF!   ANIGO   ANIGO   VAIML               VAIML   RGESG   P FLVAL	QTLSAQTLLGI.	SAAAIILILIL	IVKFRIHALLT:	LVIVSLLTALATGLPT	PAKAGTV 240 HILLIH 1111 PAKAGTV 240 HILLIH 111 PAKAGTV 240 HILLIH 111 PAKAGTV 240 HILLIH 111 HILLIH 11
35 40 45 50	orf140-1.per orf140a orf140-1.per orf140a orf140-1.per orf140a orf140-1.per orf140a orf140-1.per orf140a	P MDGWT(                 MDGWT(                 MDGWT(               P ILVKNI   GFPIF!   GFPIF!   ANIGO   ANIGO   VAIML               VAIML   RGESG   P FLVAL	QTLSAQTLLGI.	SAAAIILILIL	IVKFRIHALLT:	LVIVSLLTALATGLPT	PAKAGTV 240 HILLIH 1111 PAKAGTV 240 HILLIH 111 PAKAGTV 240 HILLIH 111 PAKAGTV 240 HILLIH 111 HILLIH 11
35 40 45 50	orf140-1.per orf140a orf140-1.per orf140a orf140-1.per orf140a orf140-1.per orf140a orf140-1.per orf140a orf140-1.per	P MDGWT(                 MDGWT(               MDGWT(               P ILVKNI  P GFPIFI   GFPIFI  ANIGQ   I           ANIGQ   VAIML   VAIML   P RGESG   I           RGESG	QTLSAQTLLGI	SAAAIILILIL	IVKFRIHALLT:	LVIVSLLTALATGLPT	PAKAGTV 240 HILLIH 1111 PAKAGTV 240 HILLIH 111 PAKAGTV 240 HILLIH 111 PAKAGTV 240 HILLIH 111 HILLIH 11
35 40 45 50	orf140-1.per orf140a orf140-1.per orf140a orf140-1.per orf140a orf140-1.per orf140a orf140-1.per orf140a orf140-1.per orf140a orf140-1.per	p MDGWT(                 MDGWT(               MDGWT(               P ILVKNI  P GFPIFI   GFPIFI  ANIGQ   I           ANIGQ  P VAIML   VAIML   P RGESG   I           RGESG	QTLSAQTLLGI	SAAAIILILIL	IVKFRIHALLT:	LVIVSLLTALATGLPT	PAKAGTV 240 HILLIH 1111 PAKAGTV 240 HILLIH 111 PAKAGTV 240 HILLIH 111 PAKAGTV 240 HILLIH 111 HILLIH 11

## Homology with a predicted ORF from N.gonorrhoeae

ORF140 shows 92% identity over a 87aa overlap with a predicted ORF (ORF140ng) from

65 N.gonorrhoeae:

```
MDGWTQTLSAQTLLGISAAAIILILILIVRFRIHALLTLVIVSLLTALATGLPTGSIVKD
                          orf140ng
                          MDGRTQTLSAQTLLGISAAAIILILILIVKFRIRALLTLVIASLLTALATGLPTGSIVND
                                                                                               60
5
           orf140.pep
                          ILVKNFGGTLGGVALLVGLGAMLERLV
                                                                                               87
                          VLVKNFGGTLGGVALLVGLGAMLGRLVETSGGAQSLADALIRMFGEKRAPFAPGVASLIF 120
           orf140ng
     The complete length ORF140ng nucleotide sequence <SEQ ID 589> was predicted to encode a
     protein having amino acid sequence <SEQ ID 590>:
                      MDGRTOTLSA OTLLGISAAA IILILILIVK FRIRALLTLV IASLLTALAT GLPTGSIVND VLVKNFGGTL GGVALLVGLG AMLGRLVETS GGAQSLADAL TRMFGEKRAP FAPGVASLIF GFPIFFDAGL IVMLPIVFAT ARRMKQDVLP FALASVGAFS VMHVFLPPHP GPIAASEFYG ANIGQVLILG LPTAFITWYF
10
                  51
                 101
                 151
                 201
                      SGYMLGKVLG RAIHVPVPEL LSGGTQDSDP PKEPAKAGTV VAVMLIPMLL
                      IFLNTGVSAL ISEKLVSADE TWVQTAKMIG STPVALLISV LAALLVLGRK
RGESGSTLEK TVDGALAPAC SVILITGAGG MFGGVLRASG IGKALADSMA
15
                 251
                 301
                      DLGIPVLLGC FLVALALRIA QGSATVALTT AAALMAPAVA AAGFTDWQLA
                 351
                      CIVLATAAGS VGCSHFNDSG FWLVGRLSDM DVPTTLKTWT VNQTLIAFIG
                 451 FALSALLFAI V*
20
     Further work revealed a variant gonococcal DNA sequence <SEQ ID 591>:
                      ATGGACGGCC GGACACAGAC GCTGTCCGCG CAAACCTTGT TGGGCATTTC
                  51
                      GGCGGCGCA ATCATCCTCA TTCTGATTTT AATCGTCAAA TTCCGCATCC
                      GCGCGCTGCT GACACTGGTC ATCGCCAGCC TGCTGACGGC TTTGGCAACC
                 101
                 151
                      GGTTTGCCCA CAGGCAGCAT CGTCAACGAC GTACTGGTCA AAAACTTCGG
25
                      CGGCACGCTC GGCGGCGTGG CGCTTCTGGT CGGTCTGGGC GCAATGCTCG
                 201
                 251
                      GACGTTTGGT AGAAACATCC GGCGGCGCAC AGTCGCTGGC GGACGCGCTG
                      ATCCGGATGT TCGGCGAAAA ACGCGCACCG TTCGCTCCGG GCGTTGCCTC
                 301
                      GCTGATTTTC GGCTTCCCGA TTTTCTTCGA TGCCGGACTA ATCGTCATGC
                 351
                      TGCCCATCGT ATTCGCCACC GCACGGCGCA TGAAACAGGA CGTACTGCCC
                 401
30
                 451
                      TTCGCGCTTG CCTCCGTCGG CGCATTTTCC GTCATGCACG TCTTCCTGCC
                 501
                      GCCCCATCCG GGCCCGATTG CCGCTTCCGA ATTTTACGGC GCGAACATCG
                 551
                      GCCAGGTTTT GATTTTGGGT CTGCCGACCG CCTTCATCAC ATGGTATTTC
                 601
                      AGCGGCTATA TGCTCGGCAA AGTGTTGGGG CGCGCCATCC ATGTTCCCGT
                 651
                      TCCCGAACTG CTCAGCGGCG GCACGCAAGA CAGCGACCCG CCGAAAGAAC
35
                      CTGCCAAAGC AGGAACGGTC GTCGCCGTCA TGCTGATTCC CATGCTGCTG
                 701
                 751
                      ATTTTCCTGA ATACCGGCGT ATCAGCCCTC ATCAGCGAAA AACTCGTAAG
                      TGCGGACGAA ACTTGGGTTC AGACGGCAAA AATGATCGGT TCGACACCTG
                 801
                 851
                      TCGCCCTTCT GATTTCCGTA TTGGCCGCAC TGTTGGTCTT GGGACGCAAA
                      CGCGGCGAAA GCGGCAGCAC GTTGGAAAAA ACCGTGGACG GCGCACTCGC
CCCCGCCTGT TCCGTGATTC TGATTACCGG CGCGGGCGGT ATGTTCGGCG
                 901
40
                 951
                1001
                      GCGTTTTGCG CGCTTCCGGC ATCGGCAAGG CACTCGCCGA CAGCATGGCG
                1051
                      GATTTGGGCA TTCCCGTCCT TTTGGGCTGC TTCCTTGTCG CCTTGGCACT
                      GCGTATCGCG CAAGGTTCGG CAACCGTCGC CCTGACCACA GCCGCCGCGC
                1101
                      TGATGGCTCC TGCCGTTGCC GCCGCCGGCT TTACCGACTG GCAGCTCGCC
                1151
45
                      TGTATCGTAT TGGCAACGGC GGCAGGTTCG GTCGGTTGCA GCCACTTCAA
                1201
                1251
                      CGACTCCGGC TTCTGGCTGG TCGGCCGCCT CTTGGATATG GACGTACCGA
                      CCACGCTGAA AACCTGGACG GTCAACCAAA CCCTCATCGC ATTCATCGGC
                1301
                1351
                      TTTGCCTTGT CCGCACTGCT GTTTGCCATC GTCTGA
      This corresponds to the amino acid sequence <SEQ ID 592; ORF140ng-1>:
50
                      MDGRTQTLSA QTLLGISAAA IILILILIVK FRIRALLTLV IASLLTALAT
                      GLPTGSIVND VLVKNFGGTL GGVALLVGLG AMLGRLVETS GGAQSLADAL IRMFGEKRAP FAPGVASLIF GFPIFFDAGL IVMLPIVFAT ARRMKQDVLP
                  51
                 101
                      FALASVGAFS VMHVFLPPHP GPIAASEFYG ANIGQVLILG LPTAFITWYF
                 151
                      SGYMLGKVLG RAIHVPVPEL LSGGTQDSDP PKEPAKAGTV VAVMLIPMLL
                 201
55
                      IFLNTGVSAL ISEKLVSADE TWVQTAKMIG STPVALLISV LAALLVLGRK
                 251
                      RGESGSTLEK TVDGALAPAC SVILITGAGG MFGGVLRASG IGKALADSMA
                      DLGIPVLLGC FLVALALRIA QGSATVALTT AAALMAPAVA AAGFTDWQLA
                 351
                      CIVLATAAGS VGCSHFNDSG FWLVGRLLDM DVPTTLKTWT VNQTLIAFIG
                 401
                      FALSALLFAI V*
60
      ORF140ng-1 and ORF140-1 show 96.3% identity over 461aa overlap:
```

WO 99/24578 PCT/IB98/01665

```
MDGWTQTLSAQTLLGISAAAIILILILIVKFRIHALLTLVIVSLLTALATGLPTGSIVND
         orf140-1
         orf140ng-1.pep VLVKNFGGTLGGVALLVGLGAMLGRLVETSGGAQSLADALIRMFGEKRAPFAPGVASLIF
                       ILVKNFGGTLGGVALLVGLGAMLGRLVETSGGAQSLADALIRMFGEKRAPFALGVASLIF
5
         orf140-1
         orf140ng-1.pep GFPIFFDAGLIVMLPIVFATARRMKQDVLPFALASVGAFSVMHVFLPPHPGPIAASEFYG
                       GFPIFFDAGLIVMLPIVFATARRMKQDVLPFALASIGAFSVMHVFLPPHPGPIAASEFYG
         orf140-1
10
         orf140ng-1.pep ANIGQVLILGLPTAFITWYFSGYMLGKVLGRAIHVPVPELLSGGTQDSDPPKEPAKAGTV
                       ANIGQVLILGLPTAFITWYFSGYMLGKVLGRTIHVPVPELLSGGTQDNDLPKEPAKAGTV
         orf140-1
         orf140ng-1.pep VAVMLIPMLLIFLNTGVSALISEKLVSADETWVQTAKMIGSTPVALLISVLAALLVLGRK
15
                       orf140-1
                       VAIMLIPMLLIFLNTGVSALISEKLVSADETWVQTAKIIGSTPIALLISVLVALFVLGRK
         orf140ng-1.pep RGESGSTLEKTVDGALAPACSVILITGAGGMFGGVLRASGIGKALADSMADLGIPVLLGC
                       20
         orf140-1
                       RGESGSALEKTVDGALAPVCSVILITGAGGMFGGVLRASGIGKALADSMADLGIPVLLGC
         orf140ng-1.pep FLVALALRIAQGSATVALTTAAALMAPAVAAAGFTDWQLACIVLATAAGSVGCSHFNDSG
                       FLVALALRIAQGSATVALTTAAALMAPAVAAAGFTDWQLACIVLATAAGSVGCSHFNDSG
25
         orf140-1
         orf140ng-1.pep FWLVGRLLDMDVPTTLKTWTVNQTLIAFIGFALSALLFAIV
                       FWLVGRLLDMDVPTTLKTWTVNQTLIALIGFALSALLFAIV
          orf140-1
     Furthermore, ORF140ng-1 is homologous to an E.coli protein:
30
          gi|882633 (U29579) ORF_o454 [Escherichia coli] >gi|1789097 (AE000358) o454;
          This 454 aa ORF is 34% identical (9 gaps) to 444 residues of an approx. 456 aa
          protein GNTP_BACLI SW: P46832 [Escherichia coli] Length = 454
           Score = 21\overline{0} bits (529), Expect = 1e-53
           Identities = 130/384 (33%), Positives = 194/384 (49%), Gaps = 19/384 (4%)
35
          Query: 88 ETSGGAQSLADALIRMFGEKRAPFAPGVASLIFGFPIFFDAGLIVMLPIVFATARRMKQD 147
                   E SGGA+SLA+
                              R G+KR A +A+
                                               G P+FFD G I++ PI++ A+ K
          Sbict: 80 EHSGGAESLANYFSRKLGDKRTIAALTLAAFFLGIPVFFDVGFIILAPIIYGFAKVAKIS 139
40
          Query: 148 VLPFALASVGAFSVMHVFLPPHPGPIAASEFYGANIGQVLILGLPTAFITWYFSGYMLGK 207
                    LFL G
                               +HV +PPHPGP+AA+
                                                A+IG + I+G+ + I
                                                                  GY
          Sbjct: 140 PLKFGLPVAGIMLTVHVAVPPHPGPVAAAGLLHADIGWLTIIGIAIS-IPVGVVGYFAAK 198
          Query: 208 VLGRAIHVPVPELL-----SGGTQDSDPPKEPAKAGTVVAVMLIPMLLIFLNTGV 257
 45
                                                   P A V ++++IP+ +I
                    ++ + +
                             E+L
                                         G T+ SD
          Sbjct: 199 IINKRQYAMSVEVLEQMQLAPASEEGATKLSDKINPPGVA-LVTSLIVIPIAIIMAGT-- 255
          Query: 258 SALISEKLVSADETWVQTAKMIGSTPXXXXXXXXXXXXXXXXXRRGESGSTLEKTVDGALA 317
                                + T ++IGS
                                                        +RG S
 50
                      +S L+
          Sbjct: 256 ---VSATLMPPSHPLLGTLQLIGSPMVALMIALVLAFWLLALRRGWSLQHTSDIMGSALP 312
          Query: 318 PACSVILITGAGGMFGGVLRASGIGKALADSMADLGIPVLLGCFLVALALRIAQGSXXXX 377
                                                        F+++LALR +QGS
                     A VIL+TGAGG+FG VL SG+GKALA+ + + +P+L
          Sbict: 313 TAAVVILVTGAGGVFGKVLVESGVGKALANMLQMIDLPLLPAAFIISLALRASQGS--AT 370
 55
          Query: 378 XXXXXXXXXXXXXXXGFTDWQLACIVLATAAGSVGCSHFNDSGFWLVGRLLDMDVPTTLK 437
                                        + LA G +G SH NDSGFW+V + L + V
                                 G
                                     0
          Sbjct: 371 VAILTTGGLLSEAVMGLNPIQCVLVTLAACFGGLGASHINDSGFWIVTKYLGLSVADGLK 430
 60
          Ouery: 438 TWTVNQTLIAFIGFALSALLFAIV 461
                    TWTV T++ F GF ++ ++A++
          Sbjct: 431 TWTVLTTILGFTGFLITWCVWAVI 454
```

Based on this analysis, including the identification of the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the

65

gonococcal protein, it is predicted that the proteins from *N. meningitidis* and *N. gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 71

The following partial DNA sequence was identified in N.meningitidis <SEO ID 593>:

```
5
                     ..GATTTCGGCA TATCGCCCGT GTATCTTTGG GTTGCCGCCG CGTTCAAACA
                 51
                       TTTGCTGTCG CCGTGGGCTG CCGACTCATA CGATGTCGCA CGCTTTGCAG
                       GCGTATTTTT TGCCGTTATC GGACTGACTT CCTGCGGCTT TGCCGGTTTC
                101
                151
                       AACTTTTTGG GCAGACACCA CGGGCGCAC. GTCGTCCTGA TTCTCATCGG
                       CTGTATCGGG CTGATTCCAG TTGCCCATTT CCTCAACCCC GCTGCCGCCG
                201
10
                251
                       CCTTTGCCGC CGCCGGACTG GTGCTGCACG GTTATTCTTT GGCTCGCCGG
                       CGCGTGATTG CCGCCTCTTT TCTGCTCGGT ACGGGCTGGA CGCTGATGTC
                301
                351
                       GTTGGCAGCA GCTTATCCGG CAGCATTTGC CCTGATGCTG CCCTTGCCCG
                401
                       TACTGATGTT TTTCCGTCCG ...
```

This corresponds to the amino acid sequence <SEQ ID 594; ORF141>:

```
1 ..DFGISPYYLW VAAAFKHLLS PWAADSYDVA RFAGVFFAVI GLTSCGFAGF
51 NFLGRHHGRX VVLILIGCIG LIPVAHFLNP AAAAFAAAGL VLHGYSLARR
101 RVIAASFLLG TGWTLMSLAA AYPAAFALML PLPVLMFFRP ...
```

Further work revealed the complete nucleotide sequence <SEQ ID 595>:

```
ATGCTGACCT ATACCCCGCC CGATGCCCGC CCGCCGCCA AAACCCACGA
20
                51
                    AAAGCCGTGG CTGCTGCTGT TGATGGCGTT TGCCTGGTTG TGGCCCGGCG
               101
                    TGTTTTCCCA CGATTTGTGG AATCCTGACG AACCTGCCGT CTATACCGCC
                    GTCGAAGCAC TGGCAGGCAG CCCCACCCC TTGGTTGCCC ATCTGTTCGG
               151
               201
                    TCAAACCGAT TTCGGCATAC CGCCCGTGTA TCTTTGGGTT GCCGCCGCGT
                    TCAAACATTT GCTGTCGCCG TGGGCTGCCG ACTCATACGA TGCCGCACGC
               251
25
               301
                    TTTGCAGGCG TATTTTTTGC CGTTATCGGA CTGACTTCCT GCGGCTTTGC
                    CGGTTTCAAC TTTTTGGGCA GACACCACGG GCGCAGCGTC GTCCTGATTC
               351
               401
                    TCATCGGCTG TATCGGGCTG ATTCCAGTTG CCCATTTCCT CAACCCCGCT
                    GCCGCCGCCT TTGCCGCCGC CGGACTGGTG CTGCACGGTT ATTCTTTGGC
               451
               501
                    TCGCCGGCGC GTGATTGCCG CCTCTTTTCT GCTCGGTACG GGCTGGACGC
30
                    TGATGTCGTT GGCAGCAGCT TATCCGGCAG CATTTGCCCT GATGCTGCCC
               551
               601
                    TTGCCCGTAC TGATGTTTTT CCGTCCGTGG CAAAGCAGGC GTTTGATGTT
                    GACGGCAGTC GCCTCACTTG CCTTTGCCCT GCCGCTTATG ACCGTTTACC
               651
               701
                    CGCTGCTCTT GGCAAAAACG CAGCCCGCGC TGTTCGCGCA ATGGCTCGAC
               751
                    TATCACGTTT TCGGTACGTT CGGCGGCGTG CGGCACGTTC AGACGGCATT
35
               801
                    CAGTTTGTTT TACTATCTGA AAAACCTGCT TTGGTTTGCA TTGCCCGCGC
                    TGCCGCTGGC GGTTTGGACG GTTTGCCGCA CGCGCCTGTT TTCGACCGAC
               851
               901
                    TGGGGGATTT TGGGCGTCGT CTGGATGCTT GCCGTTTTGG TGCTGCTTGC
               951
                    CGTCAATCCG CAGCGTTTTC AGGATAACCT CGTCTGGCTG CTTCCGCCGC
              1001
                    40
              1051
                    GCGTTTGTCA ACTGGTTCGG CATTATGGCG TTCGGACTGT TTGCCGTGTT
              1101
                    CCTGTGGACG GGCTTTTTCG CCATGAATTA CGGCTGGCCC GCCAAGCTTG
              1151
                    CCGAACGCGC CGCCTATTTC AGCCCGTATT ATGTTCCTGA TATCGATCCC
              1201
                    ATTCCGATGG CGGTTGCCGT ACTGTTCACA CCCTTGTGGC TGTGGGCGAT
                    TACCCGGAAA AACATACGCG GCAGGCAGGC GGTTACCAAC TGGGCGGCAG
              1251
45
              1301
                    GCGTTACCCT GACCTGGGCT TTGCTGATGA CGCTGTTCCT GCCGTGGCTG
              1351
                    GACGCGGCGA AAAGCCACGC GCCGGTCGTC CGGAGTATGG AGGCATCGCT
                    TTCCCCGGAA TTGAAACGGG AGCTTTCAGA CGGCATCGAG TGTATCGGCA
              1401
              1451
                    TAGGCGGCGG CGACCTGCAC ACGCGGATTG TTTGGACGCA GTACGGCACA
                    TTGCCGCACC GCGTCGGCGA TGTACAATGC CGCTACCGCA TCGTCCTCCT
               1501
50
               1551
                    GCCCCAAAAT GCGGATGCGC CGCAAGGCTG GCAGACGGTT TGGCAGGGTG
                    CGCGTCCGCG CAACAAGAC AGTAAGTTCG CACTGATACG GAAAATCGGG
               1601
                    GAAAATATAT AA
```

This corresponds to the amino acid sequence <SEQ ID 596; ORF141-1>:

	1	MLTYTPPDAR	PPAKTHEKPW	LLLLMAFAWL	WPGVFSHDLW	NPDEPAVYTA
55						WAADSYDAAR
	101	<b>FAGVFFAVIG</b>	LTSCGFAGEN	FLGRHHGRSV	VLILIGCIGL	IPVAHFLNPA
	151	AAAFAAAGLV	LHGYSLARRR	VIAASFLLGT	GWTLMSLAAA	YPAAFALMLP
	201	LPVLMFFRPW	QSRRLMLTAV	ASLAFALPLM	TVYPLLLAKT	QPALFAQWLD

-339-

```
251 YHVFGTFGGV RHVQTAFSLF YYLKNLLWFA LPALPLAVWT VCRTRLFSTD
301 WGILGVVWML AVLVLLAVNP QRFQDNLVWL LPPLALFGAA QLDSLRRGAA
351 AFVNWFGIMA FGLFAVFLWT GFFAMNYGWP AKLAERAAYF SPYYVPDIDP
401 IPMAVAVLFT PLWLWAITRK NIRGRQAVTN WAAGVTLTWA LLMTLFLPWL
5 451 DAAKSHAPVV RSMEASLSPE LKRELSDGIE CIGIGGGDLH TRIVWTQYGT
501 LPHRVGDVQC RYRIVLLPQN ADAPQGWQTV WQGARPRNKD SKFALIRKIG
551 ENI*
```

Computer analysis of this amino acid sequence gave the following results:

### Homology with a predicted ORF from N. meningitidis (strain A)

ORF141 shows 95.0% identity over a 140aa overlap with an ORF (ORF141a) from strain A of N. meningitidis:

						10	20	30
	orf141.pep				DFGISP	vylwvaaafki	ILLSPWAADSY	DVA
					1111 1	1111111111		1:1
15	orf141a				_		HLLSPWAADPY	DAA
		40	50	60	70	80	90	
			40	50	60	70	80	90
	orf141.pep	RFAGVF	AVIGLTSCG	FAGFNFLGRH	HGRXVVLILI	GCIGLIPVAH	PLNPAAAAFAA	AGL
20		11111	11:11111	<del>11</del> 1111111	111 <del>111111</del>	111111::1	Гининн	111
	orf141a	RFAGVF	AVVGLTSCG	FAGFNFLGRH	HGRSVVLILI	GCIGLIPTVH	LNPAAAAFAA	AGL
		100	110	120	130	140	150	
		•	100	110	120	130	140	
25	orf141.pep	VLHGYSI	ARRRVIAAS	FLLGTGWTLM	SLAAAYPAAF	ALMLPLPVLM		
	· • · •	111111			THITT		IIII	
	orf141a	VLHGYSI	ARRRVIAAS	FLLGTGWTLM	SLAAAYPAAF	ALMLPLPVLMI	FFRPWOSRRLM	LTA
	_	160	170	180	190	200	210	
30	orf141a	WASTAF	17 DT MTVIVDT	፣፣ አሄጥ/ ነር ነር	YOUT DOUGES	<b>ጥ</b> ኮ/~(パロロT/)ጥ?	AFSLFYYLKNL	IME
50	OTITATA	220	230	240	250	260	270	TME
		220	230	240	250	260	270	

The complete length ORF141a nucleotide sequence <SEQ ID 597> is:

	1	ATGCTGACCT	ATACCCCGCC	CGATGCCCGC	CCGCCCGCCA	AAACCCACGA
	51	AAAGCCGTGG	CTGTTGCTGT	TGATGGCGTT	TGCCTGGTTG	TGGCCCGGCG
35	101	TGTTTTCCCA	CGATTTGTGG	AATCCTGACG	AACCTGCCGT	CTATACCGCC
	151	GTCGAAGCAC	TGGCAGGCAG	CCCCACCCCT	TTGGTTGCCC	<b>ATCTGTTCGG</b>
	201	TCAAATCGAT	TTCGGCATAC	CGCCCGTGTA	TCTTTGGGTT	GCCGCCGCGT
	251	TCAAACATTT	GCTGTCGCCG	TGGGCTGCCG	ACCCGTATGA	TGCCGCACGC
	301	TTTGCCGGCG	TGTTTTTCGC	CGTTGTCGGA	CTGACTTCCT	GCGGCTTTGC
40	351	CGGTTTCAAC	TTTTTGGGCA	GACACCACGG	GCGCAGCGTC	GTCCTGATTC
	401	TCATCGGCTG	TATCGGGCTG	ATTCCGACCG	TACACTTTCT	CAACCCCGCT
	451	GCCGCCGCCT	TTGCCGCCGC	CGGACTGGTG	CTGCACGGTT	ATTCTTTGGC
	501	TCGCCGGCGC	GTGATTGCCG	CCTCTTTTCT	GCTCGGTACG	GGTTGGACGC
	551	TGATGTCGTT	GGCAGCAGCT	TATCCGGCGG	CATTTGCCCT	GATGCTGCCC
45	601	CTGCCCGTGC	TGATGTTTTT	CCGTCCGTGG	CAAAGCAGGC	GTTTGATGTT
	651	GACGGCAGTC	GCCTCGCTTG	CCTTTGCCCT	GCCGCTTATG	ACCGTTTACC
	701	CGCTGCTCTT	GGCAAAAACG	CAGCCCGCGC	TGTTCGCGCA	ATGGCTCGAC
	751	GATCACGTTT	TCGGTACGTT	CGGCGGCGTG	CGGCACATTC	AGACGGCATT
	801	CAGTTTGTTT	TACTATCTGA	AAAACCTGCT	TTGGTTTGCA	TTGCCTGCGC
50	851	TGCCGCTGGC	GGTTTGGACG	GTTTGCCGCA	CGCGCCTGTT	TTCGACCGAC
	901	TGGGGGATTT	TGGGCGTCGT	CTGGATGCTT	GCCGTTTTGG	TGCTGCTTGC
	951	CGTCAATCCG	CAGCGTTTTC	AGGATAACCT	CGTCTGGCTG	CTTCCGCCGC
	1001	TTGCCCTGTT	CGGCGCGGCG	CAACTGGACA	GCCTGAGACG	CGGCGCGGCG
	1051	GCGTTTGTCA	ACTGGTTCGG	CATTATGGCG	TTCGGACTGT	TTGCCGTGTT
55	1101	CCTGTGGACG	GGCTTTTTCG	CCATGAATTA	CGGCTGGCCC	GCCAAGCTTG
	1151	CCGAACGCGC	CGCCTATTTC	AGCCCGTATT	<b>ATGTTCCTGA</b>	TATCGATCCC
	1201	ATTCCGATGG	CGGTTGCCGT	ACTGTTCACA	CCCTTGTGGC	TGTGGGCGAT
	1251	TACCCGCAAA	AACATACGCG	GCAGGCAGGC	GGTTACCAAC	TGGGCGGCAG
	1301	GCGTTACCCT	GACCTGGGCT	TTGCTGATGA	CGCTGTTCCT	GCCGTGGCTG
60	1351	GACGCGGCGA	AAAGCCACGC	GCCCGTCGTC	CGGAGTATGG	AGGCATCGCT
	1401	TTCCCCGGAA	TTAAAACGGG	AGCTTTCAGA	CGGCATCGAG	TGTATCGACA
	1451	TAGGCGGCGG	CGACCTACAC	ACGCGGATTG	TTTGGACGCA	GTACGGCACA
	1501	TTGCCGCACC	GCGTCGGCGA	TGTACAATGC	CGCTACCGCA	TCGTCCGCTT
	1551	GCCCCAAAAC	GCGGATGCGC	CGCAAGGCTG	GCAGACGGTC	TGGCAGGGTG

```
1601 CGCGCCCGCG CAACAAGAC AGTAAGTTCG CACTGATACG GAAAACCGGG
1651 GAAAATATAT TAAAAACAAC AGATTGA
```

This encodes a protein having amino acid sequence <SEQ ID 598>:

*	1	MLTYTPPDAR	PPAKTHEKPW	LLLLMAFAWL	WPGVFSHDLW	NPDEPAVYTA
5	51	VEALAGSPTP	LVAHLFGQID	FGIPPVYLWV	AAAFKHLLSP	WAADPYDAAR
_	101	FAGVFFAVVG	LTSCGFAGFN	FLGRHHGRS <u>V</u>	VLILIGCIGL	IPTVHFLNPA
	151	AAAFAAAGLV	LHGYSLARRR	VIAASFLLGT	GWTLMSLAAA	YPAAFALMLP
	201	LPVLMFFRPW	QSRRLMLTAV	ASLAFALPLM	TVYPLLLAKT	QPALFAQWLD
	251	DHVFGTFGGV	RHIQTAFSLF	YYLKNLLWFA	LPALPLAVWT	VCRTRLFSTD
10	301	WGILGVVWML	AVLVLLAVNP	QRFQDNLVWL	LPPLALFGAA	QLDSLRRGAA
	351	AFVNWFGIMA	FGLFAVFLWT	GFFAMNYGWP	AKLAERAAYF	SPYYVPDIDP
	401			NIRGRQAVTN		
	451	DAAKSHAPVV	RSMEASLSPE	LKRELSDGIE	CIDIGGGDLH	TRIVWTQYGT
	501	LPHRVGDVQC	RYRIVRLPQN	ADAPQGWQTV	WQGARPRNKD	SKFALIRKTG
15	551	ENILKTTD*				

### ORF141a and ORF141-1 show 98.2% identity in 553 aa overlap:

	orf141a.pep	MLTYTPPDARPPAKTHEKPWLLLLMAFAWLWPGVFSHDLWNPDEPAVYTAVEALAGSPTP
20	orf141a.pep	LVAHLFGQIDFGIPPVYLWVAAAFKHLLSPWAADPYDAARFAGVFFAVVGLTSCGFAGFN
25	orf141a.pep	FLGRHHGRSVVLILIGCIGLIPTVHFLNPAAAAFAAAGLVLHGYSLARRRVIAASFLLGT
	orf141-1	FLGRHHGRSVVLILIGCIGLIPVAHFLNPAAAAFAAAGLVLHGYSLARRRVIAASFLLGT
30	orfl4la.pep	GWTLMSLAAAYPAAFALMLPLPVLMFFRPWQSRRLMLTAVASLAFALPLMTVYPLLLAKT
30	orf141-1	GWTLMSLAAAYPAAFALMLPLPVLMFFRPWQSRRLMLTAVASLAFALPLMTVYPLLLAKT
•	orf141a.pep	QPALFAQWLDDHVFGTFGGVRHIQTAFSLFYYLKNLLWFALPALPLAVWTVCRTRLFSTD
35	orf141-1	QPALFAQWLDYHVFGTFGGVRHVQTAFSLFYYLKNLLWFALPALPLAVWTVCRTRLFSTD
	orf141a.pep	WGILGVVWMLAVLVLLAVNPQRFQDNLVWLLPPLALFGAAQLDSLRRGAAAFVNWFGIMA
40	orf141-1	WGILGVVWMLAVLVLLAVNPQRFQDNLVWLLPPLALFGAAQLDSLRRGAAAFVNWFGIMA
40	orf141a.pep	FGLFAVFLWTGFFAMNYGWPAKLAERAAYFSPYYVPDIDPIPMAVAVLFTPLWLWAITRK
	orf141-1	FGLFAVFLWTGFFAMNYGWPAKLAERAAYFSPYYVPDIDPIPMAVAVLFTPLWLWAITRK
45	orf141a.pep	NIRGRQAVTNWAAGVTLTWALLMTLFLPWLDAAKSHAPVVRSMEASLSPELKRELSDGIE
	orf141-1	NIRGRQAVTNWAAGVTLTWALLMTLFLPWLDAAKSHAPVVRSMEASLSPELKRELSDGIE
50	orf141a.pep	CIDIGGGDLHTRIVWTQYGTLPHRVGDVQCRYRIVRLPQNADAPQGWQTVWQGARPRNKD
50	orf141-1	CIGIGGGDLHTRIVWTQYGTLPHRVGDVQCRYRIVLLPQNADAPQGWQTVWQGARPRNKD
	orf141a.pep	SKFALIRKTGENI
55	orf141-1	

### Homology with a predicted ORF from N.gonorrhoeae

ORF141 shows 95% identity over a 140aa overlap with a predicted ORF (ORF141ng) from

N.gonorrhoeae:

60	orf141.pep	DFGISPVYLWVAAAFKHLLSPWAADSYDVA	30
		1111 1111111111111111111111111111111111	
	orfl41ng	wnpaepavytavealagsptplvahlfgqtdfgippvylwvaaafkhllspwaahpydaa	126

	WO 99/24578	I GI/ID/5/52	
	WO 97/24570	-341-	
	orf141.pep	RFAGVFFAVIGLTSCGFAGFNFLGRHHGRXVVLILIGCIGLIPVAHFLNPAAAAFAAAGL	90
	orf141ng		186
5	orf141.pep	VLHGYSLARRRVIAASFLLGTGWTLMSLAAAYPAAFALMLPLPVLMFFRP	140
	orfl4lng		246
	An ORF141ng nu	cleotide sequence <seq 599="" id=""> was predicted to encode a protein having</seq>	amino
	acid sequence <s< th=""><th>EQ ID 600&gt;:</th><th></th></s<>	EQ ID 600>:	
10	1 !	APSEAVSARP LCEYLLHLAI RPFLLTLMLT YTPPDARPPA KTHEKPWLLL	
10	51	MAFAWLWPG VFSHDLWNPA EPAVYTAVEA LAGSPTPLVA HLFGQTDFGI	
	101	PROVINGAD FRALLSPWAD HPYDAARFAG VFFAVIGLTS CGFAGFNFLG	
	151	PHHCRSVVIT HIGCIGLIPV AHFFNPAAAA FAAAGLVLHG YSLARRRVIA	
	201	SELICTOWT IMSLAAAYPA AFALMLPLPV LMFFRPWQSR RLMLTAVASL	
15	251	AFALPIMTVY PLILAKTOPA LFAOWLNYHV FGTFGGVRHI QRAFSLFHYL	
13	201	KNLLWFAPPG LPLAVWTVCR TRLFSTDWGI LGIVWMLAVL VLLAFNPQRF	
	351	ODNLVWLLPP LALFGAAQLD SLRRGAAAFV NWFGIMAFGL FAVFLWTGFF	
		THE THE PARTY WILD TO TON AUGUS TO THE TWO TRENTS	

AMNYGWPAKL AERAAYFSPY YVPDIDPIPM AVAVLFTPLW LWAITRKNIR

GRQAVTNWAA GVTLTWALLM TLFLPWLDAA KSHAPVVRSM EASFSPELKR

ELSDGIECIG IGGGDLHTRI VWTQYGTLPH RVGDVRCRYR IVRLPQNADA

Further work revealed the following gonococcal DNA sequence <SEQ ID 601>:

551 PQGWQTVWQG ARPRNKDSKF ALIRKIGENI LKTTD\*

401

451

20

```
1 ATGCTGACCT ATACCCCGCC CGATGCCCGC CCGCCCGCCA AAACCCACGA
51 AAAACCGTGG CTGCTGCTGT TGATGGCGTT TGCCTGGCTG TGGCCCGGCG
                      TGTTTTCCCA CGATTTGTGG AATCCTGCCG AACCTGCCGT CTATACCGCC
25
                      GTCGAAGCAC TGGCAGGCAG CCCCACCCCC TTGGTTGCCC ATCTGTTCGG
                 151
                      TCAAACCGAT TTCGGCATAC CGCCCGTGTA TCTTTGGGTT GCCGCCGCAT
                      TCAAACATTT GCTGTCGCCG TGGGCAGCCG ACCCGTATGA TGCCGCACGC
                      TTTGCAGGCG TATTTTTGC CGTTATCGGA CTGACTTCTT GCGGCTTTGC
                      CGGTTTCAAC TTTTTGGGCA GACACCACGG GCGCAGCGTT GTTTTAATCC
30
                 351
                 401 ATATCGGCTG TATCGGGCTG ATTCCGGTTG CCCATTTCCT CAATCCcgcc
                      geogeoget tTGCCGCCGC CGGACTGGTG CTGCacggct actcgctgGC
                 451
                 ACGCCGGCGC GTGATtgccg cctctTtccT GCTCGGTACG GGTTGGACGT
551 TGATGTCGCT GGCGGCAGCT TATCCGGCGG CGTTTGCGCT GATGCTGCCC
                      CTGCCCGTGC TGATGTTTTT CCGTCCGTGG CAAAGCAGGC GTTTGATGTT
GACGGCAGTC GCCTCGCTTG CCTTTGCCCT GCCGCTTATG ACCGTTTACC
35
                 601
                 651
                 701 CGCTGCTCtt gGCAAAAACG CAGCCCGCGC TGTTTGCGCA ATGGCTCAAC
                  751 TATCACGTTT TCGGTACGTt cggcgGCGTG CGGCAcaTTC AGAggGCatT
                  801 Cagtttgttt cactatctgA AAaatctgct ttggttcgca ccgcccgggC
                       TGCCGCTGGC GGTTTGGACG GTTTGCCGCA CACGCCTGTT TTCGACCGAC
40
                  851
                       TGGGGGATTT TGGGCATTGT CTGGATGCTT GCCGTTTTGG TGCTGCTCGC
                  901
                       CTTTAATCCG CAGCGTTTTC AAGACAACCT CGTCTGGCTG CTGCCGCCGC
                  951
                       TTGCCCTGTT CGGCGCGGCG CAACTGGACA GCCTGAGGCG CGGCGGGGCG
                 1001
                       GCTTTTGTCA ACTGGTTCGG CATTATGGCG TTCGGGCTGT TTGCCGTGTT
                 1051
                 1101 CCTGTGGACG GGCTTTTTCG CCATGAATTA CGGCTGGCCC GCCAAGCTTG
 45
                 1151 CCGAACGCGC CGCCTACTTC AGCCCGTATT ACGTTCCCGA CATCGATCCC
                 1201 ATTCCGATGG CGGTTGCCGT ACTGTTCACA CCCTTGTGGC TGTGGGCCGAT
                       TACCCGGAAA AACATACGCG GCAGGCAGGC GGTTACCAAC TGGGCGGCAG
                 1301 GCGTTACCCT GACCTGGGCT TTGCTGATGA CGCTGTTCCT GCCGTGGCTG
                       GACGCGGCGA AAAGCCACGC GCCCGTCGTC CGGAGTATGG AGGCATCGTT
 50
                 1351
                       TTCCCCGGAA TTAAAACGGG AGCTTTCAGA CGGCATCGAG TGTATCGGCA
                 1401
                       TAGGCGGCGG CGACCTGCAC ACGCGGATTG TTTGGACGCA GTACGGCACA
                 1451
                       TTGCCGCACC GCGTCGGCGA TGTCCGTTGC CGCTACCGTA TCGTCCGCCT
                 1501
                       GCCCCAAAAC GCGGATGCGC CGCAAGGCTG GCAGACGGTC TGGCAGGGTG
                       CGCGCCCGCG CAACAAGAC AGTAAGTTTG CACTGATACG GAAAATCGGG
 55
                 1601
                       GAAAATATAT TAAAAACAAC AGATTGA
```

This corresponds to the amino acid sequence <SEQ ID 602; ORF141ng-1>:

PVYLWV AAAFKHLLSP WAADPYDAAR HHGRSV VLIHIGCIGL IPVAHFLNPA SFLLGT GWTLMSLAAA YPAAFALMLP FALPLM TVYPLLLAKT QPALFAQWLN NLLWFA PPGLPLAVWT VCRTRLFSTD DNLVWL LPPLALFGAA QLDSLRRGAA
H S F N

- 351 AFVNWFGIMA FGLFAVFLWT GFFAMNYGWP AKLAERAAYF SPYYVPDIDP 401 IPMAVAVLFT PLWLWAITRK NIRGRQAVTN WAAGVTLTWA LLMTLFLPWL 451 DAAKSHAPVV RSMEASFSPE LKRELSDGIE CIGIGGGDLH TRIVWTQYGT 501 LPHRVGDVRC RYRIVRLPQN ADAPQGWQTV WQGARPRNKD SKFALIRKIG
- 551 ENILKTTD\*

### ORF141ng-1 and ORF141-1 show 97.5% identity in 553 aa overlap:

		· · · · · · · · · · · · · · · · · · ·
	orf14lng-1.pep	${\tt MLTYTPPDARPPAKTHEKPWLLLLMAFAWLWPGVFSHDLWNPAEPAVYTAVEALAGSPTP}$
10	orf141-1	
10	orf14lng-1.pep	LVAHLFGQTDFGIPPVYLWVAAAFKHLLSPWAADPYDAARFAGVFFAVIGLTSCGFAGFN
	orf141-1	
15	orf14lng-1.pep	
	orf141-1	
20	orf14lng-1.pep	GWTLMSLAAAYPAAFALMLPLPVLMFFRPWQSRRLMLTAVASLAFALPLMTVYPLLLAKT
20	orf141-1	GWTLMSLAAAYPAAFALMLPLPVLMFFRPWQSRRLMLTAVASLAFALPLMTVYPLLLAKT
	orf141ng-1.pep	QPALFAQWLNYHVFGTFGGVRHIQRAFSLFHYLKNLLWFAPPGLPLAVWTVCRTRLFSTD
25	orf141-1	
	orf141ng-1.pep	
30	orf141-1	:
30	orfl4lng-1.pep	
	orf141-1	
35	orfl4lng-l.pep	NIRGRQAVTNWAAGVTLTWALLMTLFLPWLDAAKSHAPVVRSMEASFSPELKRELSDGIE
	orf141-1	
40	orf14lng-1.pep	
	orf141-1	
	orf14lng-1.pep	SKFALIRKIGENILKTTDX
45	orf141-1	SKFALIRKIGENIX

Based on the presence of several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 72

- 50 The following partial DNA sequence was identified in N.meningitidis <SEQ ID 603>:
  - 1 ... CAATCCGCCA AATGGTTATC GGGCCAAACT CTAGTCGGCA CAGCAATTGG
  - 51 GATACGCGGG CAGATAAAGC TTGGCGGCAA CCTGCATTAC GATATATTTA
  - 101 CCGGCCGCG ATTGAAAAAG CCCGAATTTT TCCAATCAAG GAAATGGGCA
  - 151 AGCGGTTTTC AGGTAGGCTA TACGTTTTAA
- 55 This corresponds to the amino acid sequence <SEQ ID 604; ORF142>:
  - 1 ...QSAKWLSGQT LVGTAIGIRG QIKLGGNLHY DIFTGRALKK PEFFQSRKWA
    51 SGFOVGYTF\*

Further work revealed the complete nucleotide sequence <SEQ ID 605>:

	1	ATGGATAATT	CGGGTAGTGA	GGCGACAGGA	AAATACCAAG	GAAATATCAC
	51	TTTCTCTGCC	GACAATCCTT	TGGGACTGAG	TGATATGTTC	TATGTAAATT
	101	ATGGACGTTC	GATTGGCGGT	ACGCCCGATG	AGGAAAGTTT	TGACGGCCAT
	151	CGCAAAGAAG	GCGGATCAAA	CAATTACGCC	GTACATTATT	CAGCCCCTTT
5	201	CGGTAAATGG	ACATGGGCAT	TCAATCACAA	TGGCTACCGT	TACCATCAGG
,	251	CAGTTTCCGG	ATTATCGGAA	GTCTATGACT	ATAATGGAAA	AAGTTACAAT
	301	ACTGATTTCG	GCTTCAACCG	CCTGTTGTAT	CGTGATGCCA	AACGCAAAAC
	351	CTATCTCGGT	GTAAAACTGT	GGATGAGGGA	AACAAAAAGT	TACATTGATG
	401	ATGCCGAACT	GACTGTACAA	CGGCGTAAAA	CTGCGGGTTG	GTTGGCAGAA
10	451	CTTTCCCACA	AAGAATATAT	CGGTCGCAGT	ACGGCAGATT	TTAAGTTGAA
10	501	ATATAAACGC	GGCACCGGCA	TGAAAGATGC	TCTGCGCGCG	CCTGAAGAAG
	551	CCTTTGGCGA	AGGCACGTCA	CGTATGAAAA	TTTGGACGGC	ATCGGCTGAT
	601	GTAAATACTC	CTTTTCAAAT	CGGTAAACAG	CTATTTGCCT	ATGACACATC
	651	CGTTCATGCA	CAATGGAACA	AAACCCCGCT	AACATCGCAA	GACAAACTGG
15	701	CTATCGGCGG	ACACCACACC	GTACGTGGCT	TCGACGGTGA	AATGAGTTTG
	751	TCTGCCGAGC	GGGGATGGTA	TTGGCGCAAC	GATTTGAGCT	GGCAATTTAA
	801	ACCAGGCCAT	CAGCTTTATC	TTGGGGCTGA	TGTAGGACAT	GTTTCAGGAC
	851	AATCCGCCAA	ATGGTTATCG	GGCCAAACTC	TAGTCGGCAC	AGCAATTGGG
	901	ATACGCGGGC	AGATAAAGCT	TGGCGGCAAC	CTGCATTACG	ATATATTTAC
20	951	CGGCCGCGCA	TTGAAAAAGC	CCGAATTTTT	CCAATCAAGG	AAATGGGCAA
	1001	GCGGTTTTCA	GGTAGGCTAT	ACGTTTTAA		
	_		• •	ono n	0.000.000	1.40.15
	This correspond	s to the amir	no acid seque	nce <seq ii<="" th=""><th>D 606; ORF</th><th>142-1&gt;:</th></seq>	D 606; ORF	142-1>:

This corresponds to the amino acid sequence < 5EQ ID 600; ORF 142

```
1 MDNSGSEATG KYQGNITFSA DNPLGLSDMF YVNYGRSIGG TPDEESFDGH
                51 RKEGGSNNYA VHYSAPFGKW TWAFNHNGYR YHQAVSGLSE VYDYNGKSYN
                    TDFGFNRLLY RDAKRKTYLG VKLWMRETKS YIDDAELTVQ RRKTAGWLAE
25
               101
               151 LSHKEYIGRS TADFKLKYKR GTGMKDALRA PEEAFGEGTS RMKIWTASAD
               201 VNTPFQIGKQ LFAYDTSVHA QWNKTPLTSQ DKLAIGGHHT VRGFDGEMSL
                251 SAERGWYWRN DLSWQFKPGH QLYLGADVGH VSGQSAKWLS GQTLVGTAIG
                301 IRGQIKLGGN LHYDIFTGRA LKKPEFFQSR KWASGFQVGY TF*
```

Computer analysis of this amino acid sequence gave the following results: 30

### Homology with a predicted ORF from N.gonorrhoeae

ORF142 shows 88.1% identity over a 59aa overlap with a predicted ORF (ORF142ng) from N.gonorrhoeae:

```
OSAKWLSGOTLVGTAIGIRGQIKLGGNLHY
                                                                   30
        orf142.pep
35
                                          RGWYWRNDLSWQFKPGHQLYLGADVGHVSGQSAKWLSGQTLAGTAIGIRGQIKLGGNLHY 313
        orf142ng
                  DIFTGRALKKPEFFQSRKWASGFQVGYTF
        orf142.pep
                  40
        orf142ng
                  DIFTGRALKKPEYFQTKKWVTGFQVGYSF 342
```

The complete length ORF142ng nucleotide sequence <SEQ ID 607> is:

	1	ATGGATAATT	CGGGTAGTGA	GGCGACAGGA	AAATACCAAG	GAAATATCAC
	51	TTTCTCTGCC	GACAATCCTT	TTGGACTGAG	TGATATGTTC	TATGTAAATT
	101	ATGGACGTTC	AATTGGCGGT	ACGCCCGATG	AGGAAAATTT	TGACGGCCAT
45	151	CGCAAAGAAG	GCGGATCAAA	CAATTACGCC	GTACATTATT	CAGCCCCTTT
	201	CGGTAAATGG	ACATGGGCAT	TCAATCACAA	TGGCTACCGT	TACCATCAGG
	251	CGGTTTCCGG	ATTATCGGAA	GTCTATGACT	ATAATGGAAA	AAGTTACAAC
	301	ACTGATTTCG	GCTTCAACCG	CCTGTTGTAT	CGTGATGCCA	AACGCAAAAC
	351	CTATCTCAGT	GTAAAACTGT	GGACGAGGGA	AACAAAAAGT	TACATTGATG
50	401	ATGCCGAACT	GACTGTACAA	CGGCGTAAAA	CCACAGGTTG	GTTGGCAGAA
	451	CTTTCCCACA	AAGGATATAT	CGGTCGCAGT	ACGGCAGATT	TTAAGTTGAA
	501	ATATAAACAC	GGCACCGGCA	TGAAAGATGC	TCTGCGCGCG	CCTGAAGAAG
	551	CCTTTGGCGA	AGGCACGTCA	CGTATGAAAA	TTTGGACGGC	ATCGGCTGAT
	601	GTAAATACTC	CTTTTCAAAT	CGGTAAACAG	CTATTTGCCT	ATGACACATC
55	651	CGTTCATGCA	CAATGGAACA	AAACCCCGCT	AACATCGCAA	GACAAACTGG
	701	CTATCGGCGG	ACACCACACC	GTACGTGGCT	TCGACGGTGA	AATGAGTTTG
	751	CCTGCCGAGC	GGGGATGGTA	TTGGCGCAAC	GATTTGAGCT	GGCAATTTAA
	801	ACCAGGCCAT	CAGCTTTATC	TTGGGGCTGA	TGTAGGACAT	GTTTCAGGAC
	851	AATCCGCCAA	ATGGTTATCG	GGCCAAACTC	TAGCCGGCAC	AGCAATTGGG
60	901	ATACGCGGGC	AGATAAAGCT	TGGCGGCAAC	CTGCATTACG	ATATATTTAC
•	951	CGGCCGTGCA	TTGAAAAAGC	CCGAATATTT	TCAGACGAAG	AAATGGGTAA

#### 1001 CGGGGTTTCA GGTGGGTTAT TCGTTTTGA

This encodes a protein having amino acid sequence <SEQ ID 608>:

```
1 MDNSGSEATG KYQGNITFSA DNPFGLSDMF YVNYGRSIGG TPDEENFDGH
51 RKEGGSNNYA VHYSAPFGKW TWAFNHNGYR YHQAVSGLSE VYDYNGKSYN
5 101 TDFGFNRLLY RDAKRKTYLS VKLWTRETKS YIDDAELTVQ RRKTTGWLAE
151 LSHKGYIGRS TADFKLKYKH GTGMKDALRA PEEAFGEGTS RMKIWTASAD
201 VNTPFQIGKQ LFAYDTSVHA QWNKTPLTSQ DKLAIGGHHT VRGFDGEMSL
251 PAERGWYWRN DLSWQFKPGH QLYLGADVGH VSGQSAKWLS GQTLAGTAIG
301 IRGQIKLGGN LHYDIFTGRA LKKPEYFQTK KWVTGFQVGY SF*
```

The underlined sequence (aromatic-Xaa-aromatic amino acid motif) is usually found at the C-terminal end of outer membrane proteins.

### ORF142ng and ORF142-1 show 95.6% identity over 342aa overlap:

```
orf142-1.pep MDNSGSEATGKYQGNITFSADNPLGLSDMFYVNYGRSIGGTPDEESFDGHRKEGGSNNYA
                 15
       orf142ng-1
                 MDNSGSEATGKYQGNITFSADNPFGLSDMFYVNYGRSIGGTPDEENFDGHRKEGGSNNYA
       orf142-1.pep VHYSAPFGKWTWAFNHNGYRYHQAVSGLSEVYDYNGKSYNTDFGFNRLLYRDAKRKTYLG
                 ####||
       orf142ng-1
                 VHYSAPFGKWTWAFNHNGYRYHQAVSGLSEVYDYNGKSYNTDFGFNRLLYRDAKRKTYLS
20
       orf142-1.pep VKLWMRETKSYIDDAELTVQRRKTAGWLAELSHKEYIGRSTADFKLKYKRGTGMKDALRA
                 orf142ng-1
                 VKLWTRETKSYIDDAELTVQRRKTTGWLAELSHKGYIGRSTADFKLKYKHGTGMKDALRA
25
       orf142-1.pep PEEAFGEGTSRMKIWTASADVNTPFQIGKQLFAYDTSVHAQWNKTPLTSODKLAIGGHHT
                 orf142ng-1
                 PEEAFGEGTSRMKIWTASADVNTPFQIGKQLFAYDTSVHAQWNKTPLTSODKLAIGGHHT
       orf142-1.pep VRGFDGEMSLSAERGWYWRNDLSWQFKPGHQLYLGADVGHVSGQSAKWLSGOTLVGTAIG
30
                 orf142ng-1
                 VRGFDGEMSLPAERGWYWRNDLSWQFKPGHQLYLGADVGHVSGQSAKWLSGQTLAGTAIG
        orf142-1.pep_IRGQIKLGGNLHYDIFTGRALKKPEFFQSRKWASGFQVGYTF
                 35
        orf142ng-1
                 IRGQIKLGGNLHYDIFTGRALKKPEYFQTKKWVTGFQVGYSF
```

#### In addition, ORF142ng is homologous to the HecB protein of *E. chrysanthemi*:

```
gi|1772622 (L39897) HecB [Erwinia chrysanthemi] Length = 558
           Score = 119 bits (295), Expect = 3e-26
           Identities = 88/346 (25%), Positives = 151/346 (43%), Gaps = 22/346 (6%)
40
                     DNSGSEATGKYQGNITFSADNPFGLSDMFYVNYGRSIGGTPDEENFDGHRKEGGSNNYAV 61
                     DNSG ++TG+ Q N + + DN FGL+D ++++ G S +
                                                               + D
          Sbjct: 230 DNSGQKSTGEEQLNGSLALDNVFGLADQWFISAGHS---SRFATSHDAESLQAG----- 280
45
          Query: 62 HYSAPFGKWTWAFNHNGYRYHQAVSGLSEVYDYNGKSYNTDFGFNRLLYRDAKRKTYLSV 121
                      +S P+G W
                                +N++ RY
                                                      GS
                                                             F +R+++RD
          Sbjct: 281 -FSMPYGYWNLGYNYSQSRYRNTFINRDFPWHSTGDSDTHRFSLSRVVFRDGTMKTAIAG 339
          Query: 122 KLWTRETKSYIDDAELTVQRRKTTGWLAELSHKGYIGRSTADFKLKYKHGTGMKDALRAP 181
50
                             +Y+++L
                                        RK +
                                                 ++H
                                                            A F
                                                                  ΥG
          Sbjct: 340 TFSQRTGNNYLNGSLLPSSSRKLSSVSLGVNHSQKLWGGLATFNPTYNRGVRWLGSETDT 399
          Query: 182 EEAFGEGTSRMKIWTASADVNTPFQIGKQLFAYDTSVHAQWNKTPLTSQDKLAIGGHHTV 241
                     +++ E +
                                  WT SA
                                          P
                                                    Y S++ Q++ L
55
          Sbjct: 400 DKSADEPRAEFNKWTLSASYYHPV---TDSITYLGSLYGQYSARALYGSEOLTLGGESSI 456
          Query: 242 RGFDGEMSLPAERGWYWRNDLSWQFKP----GHQLYLGA-DVGHVSGQSAKWLSGQTLAG 296
                     RGF E
                                 RG YWRN+L+WQ
                                                   G+ ++ A D GH+
                                                                         + +L G
           Sbjct: 457 RGF-REQYTSGNRGAYWRNELNWQAWQLPVLGNVTFMAAVDGGHLYNHKQDNSTAASLWG 515
60 ·
           Query: 297 TAIGIRGQIKLGGNLHYDIFTGRALKKPEYFQTKKWVTGFQVGYSF 342
```

+ G + P + Q

V G++VG SF

L

A+G+

Sbjct: 516 GAVGMTVASRW---LSQQVTVGWPISYPAWLQPDTMVVGYRVGLSF 558

On the basis of this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### 5 Example 73

The following partial DNA sequence was identified in N.meningitidis <SEQ ID 609>:

```
1 ATGCGGACGA AATGGTCAGC AGTGAGAAGC TGCTTACTTG GGCGGACACC
51 GCCGACATCG ATACCGCTTT GAACCTGTTG TACCGTTTGC AAAAACTCGA
101 ATTCCTCTAT GGCGATGAAA ACGGTCATTC AGACGGCATC AATTTGWCGG
151 ACGAGCAATT GCCGTTGCTG ATGGAACAAT TGTCCGGCAG CGGTAAGGCG
201 TTATTGGTCG ATCGGAACGG TCTGTATCTT GCCACACGCCA ATTTCCATCA
251 TGAGGCGGCG GAAGAGTTGG GGTTGTTGGC GGCAGAAGTC GCACAGATGG
301 AAAAGAAATA CCGGCTGCTG ATTAAGAACA AC..
```

This corresponds to the amino acid sequence <SEQ ID 610; ORF143>:

```
15 MRTKWSAVRS CTWADTADID TALNLLYRLQ KLEFLYGDEN GHSDGINLXD
51 EQLPLLMEQL SGSGKALLVD RNGLYLANAN FHHEAAEELG LLAAEVAQME
101 KKYRLLIKNN ..
```

Further work revealed the complete nucleotide sequence <SEQ ID 611>:

```
ATGGAATCAA CACTTTCACT ACAAGCAAAT TTATATCCCC GCCTGACTCC
                    TGCCGGTGCA TTTTATGCCG TATCCAGCGA TGCCCCCAGT GCCGGTAAAA
20
                    CTTTGTTGCA CAGCCTGTTG AAAGCAGATG CGGACGAAAT GGTCAGCAGT
               101
                    GAGAAGCTGC TTACTTGGGC GGACACCGCC GACATCGATA CCGCTTTGAA
                201 CCTGTTGTAC CGTTTGCAAA AACTCGAATT CCTCTATGGC GATGAAAACG
                251 GTCATTCAGA CGGCATCAAT TTGTCGGACG AGCAATTGCC GTTGCTGATG
                    GAACAATTGT CCGGCAGCGG TAAGGCGTTA TTGGTCGATC GGAACGGTCT
25
                301
                351 GTATCTTGCC AACGCCAATT TCCATCATGA GGCGGCGGAA GAGTTGGGGT
                    TGTTGGCGGC AGAAGTCGCA CAGATGGAAA AGAAATACCG GCTGCTGATT
                401
                451 AAGAACAACC TGTATATCAA CAATAACGCT TGGGGCGTTT GCGATCCTTC
                     CGGTCAGAGC GAATTGACAT TTTTCCCATT GTATATCGGT TCAACCAAAT
                501
                     TTATTTTGGT TATCGGCGGC ATTCCCGATT TGGGCAAAGA GGCATTTGTT
30
                551
                601 ACTITGGTAA GGATTTTATA CCGCCGTTAC AGCAACCGCG TGTAA
```

This corresponds to the amino acid sequence <SEQ ID 612; ORF143-1>:

```
35 MESTLSLQAN LYPRLTPAGA FYAVSSDAPS AGKTLLHSLL KADADEMVSS
EKLLTWADTA DIDTALNLLY RLQKLEFLYG DENGHSDGIN LSDEQLPLLM
EQLSGSGKAL LVDRNGLYLA NANFHHEAAE ELGLLAAEVA QMEKKYRLLI
151 KNNLYINNNA WGVCDPSGQS ELTFFPLYIG STKFILVIGG IPDLGKEAFV
201 TLVRILYRRY SNRV*
```

Computer analysis of this amino acid sequence gave the following results:

# Homology with a predicted ORF from N. meningitidis (strain A)

ORF143 shows 92.4% identity over a 105aa overlap with an ORF (ORF143a) from strain A of N. meningitidis:

					10	20 . השת הדהשתו	30 NLLYRLQKLEFL
	orf143.pep				1:: 11	1 1111111	1111111111
45	orf143a	GAFYAVSS	DXPSAGKTLL	HSLLKADADE	MVSSEKLLTW	AXTADIDTAL	NLLYRLQKLEFL
45	<b>4-</b>	20	30	40	50	60	70
		40	50	60	70	80	90
•	orf143.pep	YGDENGHS	DGINLXDEQI	PLLMEQLSGS	GKALLVDRNG	LYLANANFHH	EAAEELGLLAAE
50		11111111	11111 1111	шшш	111111111	[ ] ] ] ] ] ] ] ]	шшшш

acid sequence <SEQ ID 616>:

-346-

		YGDENGHSDGINLSDEQLPLLMEQLSGSGKALLVDRNGLYLANANFHHEAAEELGLLAAE 80 90 100 110 120 130	
5		100 110 VAQMEKKYRLLIKNN                 VAQMEKKYRLXIKNNLYINNNAWGVCDPSGQSELTFFPLYIGSTKFILVIGGIPDLGKEA	
		140 150 160 170 180 190	
	The complete length	ORF143a nucleotide sequence <seq 613="" id=""> is:</seq>	
10	51 TGC 101 CTT 151 GAG	GGAATCAA CANTTTCACT ACAAGCAAAT TTATATCNCC GCCTGACTCC CCGGTGCA TTTTATGCCG TATCCAGCGA TGNCCCCAGT GCCGGTAAAA TTGTTGCA CAGCCTGTTG AAAGCGGATG CGGACGAAAT GGTNAGCAGT GAAGCTGC TTACCTGGGC GGANACCGCC GACATCGATA CCGCTTTGAA	
15	251 GTC 301 GAA 351 GTA 401 TGT	TETTGTAC CETTTGCAAA AACTCGAATT CCTCTATGGC GATGAAAACG CATTCAGA CGCCATCAAT TTGTCGGACG AGCAATTGCC GTTGCTGATG ACAATTGT CCGGCAGCGG TAAGGCGTTA TTGGTCGATC GGAACGGTCT ATCTTGCC AACGCCAATT TCCATCATGA GGCGGCGGAA GAGTTGGGGT TTGGCGGC AGAAGTCGCA CAGATGGAAA AGAAATACCG GCTGCNNATT	
20	501 CGG 551 TTA 601 ACT	GAACAACC TGTATATCAA CAATAACGCT TGGGGCGTTT GCGATCCTTC GTCAGAGC GAATTGACAT TTTTCCCATT GTATATCGGT TCAACCAAAT ATTTTGGT TATCGGCGGC ATTCCCGATT TGGGCAAAGA GGCATTTGTT GTTGGTAA GGATNTTATA CCNCCNGTTA CAGCAACCGC GTGTAAAACT GGAGAGAG GANGGGTTAT GCAGCAATTA TTGA	
	This encodes a prote	ein having amino acid sequence <seq 614="" id="">:</seq>	
25	51 EKL 101 EQL 151 KNN	STXSLQAN LYXRLTPAGA FYAVSSDXPS AGKTLLHSLL KADADEMVSS LITWAXTA DIDTALNLLY RLQKLEFLYG DENGHSDGIN LSDEQLPLLM LSGSGKAL LVDRNGLYLA NANFHHEAAE ELGLLAAEVA QMEKKYRLXI NLYINNNA WGVCDPSGQS ELTFFPLYIG STKFILVIGG IPDLGKEAFV VRXLYXXL QQPRVKLGRE XGLCSNY*	<b>i</b>
30	ORF143a and ORF1	143-1 show 97.1% identity in 207 aa overlap:	
	orf143a.pep	MESTXSLQANLYXRLTPAGAFYAVSSDXPSAGKTLLHSLLKADADEMVSSEKLLTWAXTA	
35	orf143a.pep	DIDTALNLLYRLQKLEFLYGDENGHSDGINLSDEQLPLLMEQLSGSGKALLVDRNGLYLA	
40	orf143a.pep orf143-1	NANFHHEAAEELGLLAAEVAQMEKKYRLXIKNNLYINNNAWGVCDPSGQSELTFFPLYIG 	
45	orf143a.pep orf143-1	STKFILVIGGIPDLGKEAFVTLVRXLY	
	Homology with a pr	redicted ORF from N. gonorrhoeae	
	ORF143 shows 95.5	5% identity over a 110aa overlap with a predicted ORF (ORF143ng)	from
	N.gonorrhoeae:		
50		MRTKWSAVRSCTWADTADIDTALNLLYRLQKLEFLYGDENGHSDGINLXDEQLPLLMEQL	60 60
55	orf143.pep orf143ng	111111111111111111111111111111111111111	110
	An ORF143ng nucle	eotide sequence <seq 615="" id=""> was predicted to encode a protein having a</seq>	ımino

-347-

```
1 MRTKWSAVRS CSRADTADID TALNLLYRLQ KLEFLYGDEN GHSDGINLSD
               51 EQLPLLMEQL SGSGKALLVD RNGLYLANAN FHHESAEELG LLAAEVAQME
              101 KKYRLLIRNN LYINNNAWGV CDPSGQSELT FFPLYIGSTK FILVIAGIPD
                    LSKGGICYFG KDFIPPLQQP RVKLGTGGIM RQLLISILED LNNTSTDIIA
               151
                    SAVISTOGLP MATMLPSHLN SDRVGAISAT LLALGSRSVQ ELACGELEQV
               251 MIKGKSGYIL LSQAGKDAVL VLVAKETGRL GLILLDAKRA ARHIAEAI*
5
    Further work revealed the following gonococcal DNA sequence <SEQ ID 617>:
                 1 ATGGAATCAA CACTTTCACT ACAAGCGAAT TTATATCCCT GCCTGACTCC
                    TGCCGGTGCA TTTTATGCCG TATCCAGCGA TGCCCCCAGT GCCGGTAAAA
                51
                    CTTTGTTGCG CAGCCTGTTG AAAGCGGATG CGGACGAAGT GGTCAGCAGT
10
               101
                    GAGAAGCTGC TCGCGGCGGA CACCGCCGAC ATCGATACCG CTTTGAACCT
               151
               201 GTTGTACCGT TTGCAAAAAC TCGAATTCCT CTATGGCGAT GAAAACGGTC
               251 ATTCAGACGG CATCAATTTG TCGGACGAGC AATTGCCGTT GCTGATGGAA
                    CAATTGTCCG GCAGCGGTAA GGCATTATTG GTCGATCGGA ACGGTCTGTA
               301
                    TCTTGCCAAC GCCAATTTCC ATCATGAGTC GGCGGAAGAG TTGGGGTTGT
               351
                    TGGCGGCAGA AGTCGCACAG ATGGAAAAGA AATACCGGCT GCTGATTAGG
15
                401
                    AACAACCTGT ATATCAACAA TAACGCTTGG GGCGTTTGCG ATCCTTCCGG
                    TCAGAGCGAA TTGACATTTT TCCCATTGTA TATCGGTTCA ACCAAATTTA
                451
                    TTTTGGTTAT CGCCGGCATT CCCGATTTGA GCAAAGAGGC ATTTGTTACT
                501
                551
                    TTGGTAAGGA TTTTATACCG CCGTTACAGC AACCGCGTGT AA
                601
20
     This corresponds to the amino acid sequence <SEQ ID 618; ORF143ng-1>:
                  1 MESTLSLQAN LYPCLTPAGA FYAVSSDAPS AGKTLLRSLL KADADEVVSS
                    EKLLAADTAD IDTALNLLYR LQKLEFLYGD ENGHSDGINL SDEQLPLLME
                 51
                     QLSGSGKALL VDRNGLYLAN ANFHHESAEE LGLLAAEVAQ MEKKYRLLIR
                     NNLYINNNAW GVCDPSGQSE LTFFPLYIGS TKFILVIAGI PDLSKEAFVT
25
                151
                201 LVRILYRRYS NRV*
      ORF143ng-1 and ORF143-1 show 95.8% identity in 214 aa overlap:
           orf143ng-1.pep MESTLSLQANLYPCLTPAGAFYAVSSDAPSAGKTLLRSLLKADADEVVSSEKLLA-ADTA
                          առանում առանանանում անանանանում ան
                          MESTLSLQANLYPRLTPAGAFYAVSSDAPSAGKTLLHSLLKADADEMVSSEKLLTWADTA
            orf143-1
 30
            orfl43ng-1.pep DIDTALNLLYRLQKLEFLYGDENGHSDGINLSDEQLPLLMEQLSGSGKALLVDRNGLYLA 119
                           DIDTALNLLYRLOKLEFLYGDENGHSDGINLSDEQLPLLMEQLSGSGKALLVDRNGLYLA 120
            orf143-1
            orf143ng-1.pep NANFHHESAEELGLLAAEVAQMEKKYRLLIRNNLYINNNAWGVCDPSGQSELTFFPLYIG 179
 35
                           idanii: Arrigia anti                            NANFHHEAAEELGLLAAEVAQMEKKYRLLIKNNLYINNNAWGVCDPSGQSELTFFPLYIG 180
            orf143-1
```

Based on the presence of the putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 74

40

45

The following partial DNA sequence was identified in N.meningitidis <SEQ ID 619>:

```
1 ATGACCTTTT TACAACGTTT GCAAGGTTTG GCAGACAATA AAATCTGTGC
                    GTTTGCATGG TTCGTCGTCC GCCGCTTTGA TGAAGAACGC GTACCGCAGI
                51
                    CGGCGGCAAG CATGACGTTT ACGACGCTGC TGGCACTCGT CCCCGTGCTG
               101
50
                    ACCGTGATGG TGGCGGTCGC TTCGATTTTC CCCGTGTTCG ACCGCTGGTC
               151
                    GGATTCGTTC GTCTCCTTCG TCAACCAAAC CATTGTGCCG CA.GGCGCGG
               201
               251 ACATGGTGTT CGACTATATC AATGCGTTCC GCGAGCAGGC GAACCGGCTG
                    ACGGCAATCG GCAGCGTGAT GCTGGTCGTT ACCTCGCTGA TGCTGATTCG
               301
               351 GACGATAGAC AATACGTTCA ACCGCATCTG GACGGGTCAA WTYCCAGCGT
55
                401 CCGTGGATG...
```

orf143ng-1.pep STKFILVIAGIPDLSKEAFVTLVRILYRRYSNRV

This corresponds to the amino acid sequence <SEQ ID 620; ORF144>:

```
1 MTFLQRLQGL ADNKICAFAW FVVRRFDEER VPQXAASMTF TTLLALVPVL
51 TVMVAVASIF PVFDRWSDSF VSFVNQTIVP XGADMVFDYI NAFREQANRL
101 TAIGSVMLVV TSLMLIRTID NTFNRIWRVX XQRPWM...
```

5 Further work revealed the complete nucleotide sequence <SEQ ID 621>:

```
ATGACCTTTT TACAACGTTT GCAAGGTTTG GCAGACAATA AAATCTGTGC
                51
                    GTTTGCATGG TTCGTCGTCC GCCGCTTTGA TGAAGAACGC GTACCGCAGG
               101
                    CGGCGGCAAG CATGACGTTT ACGACGCTGC TGGCACTCGT CCCCGTGCTG
                    ACCGTGATGG TGGCGGTCGC TTCGATTTTC CCCGTGTTCG ACCGCTGGTC
               151
10
                201
                     GGATTCGTTC GTCTCCTTCG TCAACCAAAC CATTGTGCCG CAGGGCGCGG
                    ACATGGTGTT CGACTATATC AATGCGTTCC GCGAGCAGGC GAACCGGCTG
                251
                301
                     ACGGCAATCG GCAGCGTGAT GCTGGTCGTT ACCTCGCTGA TGCTGATTCG
                351
                     GACGATAGAC AATACGTTCA ACCGCATCTG GCGGGTCAAT TCCCAGCGTC
                401
                     CGTGGATGAT GCAGTTTCTC GTCTATTGGG CTTTACTGAC GTTCGGGCCG
15
                451
                     CTGTCTTTGG GCGTGGGCAT TTCCTTTATG GTCGGCTCGG TACAGGATGC
                501
                     CGCGCTTGCC TCAGGTGCGC CGCAGTGGTC GGGCGCGTTG CGAACGGCGG
                551
                    CGACGCTGAC CTTCATGACG CTTTTGCTGT GGGGGCTGTA CCGCTTCGTG
                601
                     CCAAACCGCT TCGTTCCCGC GCGGCAGGCG TTTGTCGGGG CTTTGGCAAC
                651
                     AGCGTTTTGT CTGGAAACCG CGCGCTCCCT CTTCACTTGG TATATGGGCA
20
                     ATTTCGACGG CTACCGCTCG ATTTACGGCG CGTTTGCCGC CGTGCCGTTT
                701
                751
                     TTTCTGTTGT GGCTGAACCT GTTGTGGACG CTGGTCTTGG GCGGCGCGGT
                801
                    GCTGACTTCT TCACTCTCCT ACTGGCAGGG AGAAGCGTTC CGCAGGGGCT
                851
                     TCGACTCGCG CGGACGGTTT GACGACGTGT TGAAAATCCT GCTGCTTCTG
                901
                     GATGCGGCGC AAAAAGAAGG CAAAGCCTTG CCTGTTCAGG AGTTCAGACG
25
                951
                     GCATATCAAT ATGGGCTACG ACGAGTTGGG CGAGCTTTTG GAAAAGCTGG
               1001
                     CGCGGCACGG CTACATCTAT TCCGGCAGAC AGGGTTGGGT GTTGAAAACG
                     GGGGCGGATT CGATTGAGTT GAACGAACTC TTCAAGCTCT TCGTTTACCG
               1051
               1101
                     TCCGTTGCCT GTGGAAAGGG ATCATGTGAA CCAAGCTGTC GATGCGGTAA
               1151
                     TGACACCGTG TTTGCAGACT TTGAACATGA CGCTGGCAGA GTTTGACGCT
30
               1201
                     CAGGCGAAAA AACGGCAGTA G
```

This corresponds to the amino acid sequence <SEQ ID 622; ORF144-1>:

```
MTFLQRLQGL ADNKICAFAW FVVRRFDEER VPQAAASMTF TTLLALVPVL
                     TVMVAVASIF PVFDRWSDSF VSFVNQTIVP QGADMVFDYI NAFREQANRL
                     TAIGSVMLVV TSLMLIRTID NTFNRIWRVN SQRPWMMQFL VYWALLTFGP
               101
35
                    LSLGVGISFM VGSVQDAALA SGAPQWSGAL RTAATLTFMT LLLWGLYRFV
               151
               201
                    PNRFVPARQA FVGALATAFC LETARSLETW YMGNFDGYRS IYGAFAAVPF
               251
                     FLLWLNLLWT LVLGGAVLTS SLSYWQGEAF RRGFDSRGRF DDVLKILLLL
                301
                    DAAQKEGKAL PVQEFRRHIN MGYDELGELL EKLARHGYIY SGRQGWVLKT
                351
                    GADSIELNEL FKLFVYRPLP VERDHVNQAV DAVMTPCLQT LNMTLAEFDA
40
                401
                    QAKKRQ*
```

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from N. meningitidis (strain A)

ORF144 shows 96.3% identity over a 136aa overlap with an ORF (ORF144a) from strain A of N. meningitidis:

45		10	20	30	40	50	60
	orf144.pep	MTFLQRLQGLADNKIC					
	•	<u> </u>	<del>111</del> 1111	111111111	1	111111111	$\overline{\Pi}\overline{\Pi}$
	orfl44a	MTFLQRLQGLADNKIC		RRFDEERVPQA	aasmtftt <u>li</u>	ALVPVLTVMV	AVASIF
50		10	20	30	40	50	60
50							
	C1 4 4	70	80	90	100	110	120
	orf144.pep	PVFDRWSDSFVSFVNC	ZTIVPXGA	DMVFDYINAFR	EQANRLTAIC	SVMLVVTSLM	LIRTID
,	orf144a	DAEDDRODGERGERAIC	MITUDOCA			11111111111	111111
55	OLLIANA	PVFDRWSDSFVSFVNC	80 80 XIIV	90	100		
<i>JJ</i>		70	80	. 30	100	110	120
		130					
	orfl44.pep	NTFNRIWRVXXQRPWN	§ .				
		111111111 11111					
60	orf144a	NTFNRIWRVNSQRPWN	MQFLVYW	ALLTFGPLSLG	<u>VGISFXV</u> GS\	/QDAALASGAP	QWSGAL

-349-

180 160 170 140 150 130 The complete length ORF144a nucleotide sequence <SEQ ID 623> is: 1 ATGACCTTTT TACAACGTTT GCAAGGTTTG GCAGACAATA AAATCTGTGC 51 GTTTGCATGG TTCGTCGTCC GCCGCTTTGA TGAAGAACGC GTACCGCAGG CGGCGGCAAG CATGACGTTT ACGACACTGC TGGCACTCGT CCCCGTGCTG 5 151 ACCGTGATGG TGGCGGTCGC TTCGATTTTC CCCGTGTTCG ACCGNTGGTC 101 GGATTCGTTC GTCTCCTTCG TCAACCAAAC CATTGTGCCG CAGGGCGCGG 201 ACATGGTNTT CGACTATATC AATGCGTTCC GCGAGCAGGC GAACCGGCTG 251 ACGGCAATCG GCAGCGTGAT GCTGGTCGTT ACCTCGCNGA TGCTGATTCG 301 GACGATAGAC AATACGTTCA ACCGCATCTG GCGGGTCAAT TCCCAGCGTC 10 351 CGTGGATGAT GCAGTTTCTC GTCTATTGGG CTTTACTGAC GTTCGGGCCG 401 CTGTCTTTGG GCGTGGGCAT TTCCTTTATN GTCGGCTCGG TACAGGATGC 451 CGCGCTTGCC TCAGGTGCGC CGCAGTGGTC GGGCGCGTTG CGAACGGCGG 501 CGACGCTGAN CTTCATGACG CTTTTGCTGT GGGGGCTGTA CCGCTNCGTG 551 CCAAACCGCT TCGTTCCCGC GCGGCANGCG TTTGTCGGGG CTTTGGCAAC 601 15 651 AGCGTTCTGT CTGGAAACCG CGCGTTCCCT CTTTACTTGG TATATGGGCA 701 ATTTCGACGG CTACCGCTCG ATTTACGGNG CGTTTGCCGC CGTGCCGTTT TTTCTGTTGT GGCTGAACCT GTTGTGGACG CTGGTCTTGG GCGGCGCGGT 751 GCTGACTTCT TCACTCTCCT ACTGGCAGGG AGAAGCGTTC CGCAGGGNCT 801 TCGACTCGCG CGGACGGTTT GACGACGTGT TGAAAATCCT GCTGCTTCTG 851 20 GATGCGGCGC AAAAAGAAGG CNAAGCCTTG CCTGTTCAGG AGTTCAGACG 901 GCATATCAAT ATGGGCTACG ACGAGTTGGG CGAGCTTTTG GAAAAGCTGG 951 CGCGGCACGG CTACATCTAT TCCGGCAGAC AGGGTTGGGT GTTGAAAACG GGGGCGGATT CGATTGAGTT GAACGAACTC TTCAAGCTCT TCGTTTACCG 1001 1051 TCCGTTGCCT GTGGAAAGGG ATCATGTGAA CCAAGCTGTC GATGCGGTAA 25 1101 TGATGCCGTG TTTGCAGACT TTGAACATGA CGCTGGCAGA GTTTGACGCT 1151 CAGGCGAAAA AACAGCAGCA ATCTTGA 1201 This encodes a protein having amino acid sequence <SEQ ID 624>: MTFLQRLQGL ADNKICAFAW FVVRRFDEER VPQAAASMTF TTLLALVPVL TVMVAVASIF PVFDRWSDSF VSFVNQTIVP QGADMVFDYI NAFREQANRL TAIGSVMLVV TSXMLIRTID NTFNRIWRVN SQRPWMMQFL VYWALLTFGP 30 51 101 LSLGVGISFX VGSVQDAALA SGAPQWSGAL RTAATLXFMT LLLWGLYRXV 151 PNRFVPARXA FVGALATAFC LETARSLETW YMGNFDGYRS IYGAFAAVPF FLLWLNLLWT LVLGGAVLTS SLSYWQGEAF RRXFDSRGRF DDVLKILLLL 201 251 DAAQKEGXAL PVQEFRRHIN MGYDELGELL EKLARHGYIY SGRQGWVLKT 35 301 GADSIELNEL FKLFVYRPLP VERDHVNQAV DAVMMPCLQT LNMTLAEFDA 351 QAKKQQQS\* 401 ORF144a and ORF144-1 show 97.8% identity in 406 aa overlap: MTFLQRLQGLADNKICAFAWFVVRRFDEERVPQAAASMTFTTLLALVPVLTVMVAVASIF orf144a.pep 40 MTFLQRLQGLADNKICAFAWFVVRRFDEERVPQAAASMTFTTLLALVPVLTVMVAVASIF orf144-1 PVFDRWSDSFVSFVNQTIVPQGADMVFDYINAFREQANRLTAIGSVMLVVTSXMLIRTID orf144a.pep 1996 - 1996 - 1996 - 1996 - 1996 - 1996 - 1996 - 1996 - 1996 - 1996 - 1996 - 1996 - 1996 - 1996 - 1996 - 1996 PVFDRWSDSFVSFVNQTIVPQGADMVFDYINAFREQANRLTAIGSVMLVVTSLMLIRTID 45 orf144-1 NTFNRIWRVNSQRPWMMQFLVYWALLTFGPLSLGVGISFXVGSVQDAALASGAPQWSGAL orf144a.pep NTFNRIWRVNSQRPWMMQFLVYWALLTFGPLSLGVGISFMVGSVQDAALASGAPQWSGAL orf144-1 50 RTAATLXFMTLLLWGLYRXVPNRFVPARXAFVGALATAFCLETARSLFTWYMGNFDGYRS orf144a.pep RTAATLTFMTLLLWGLYRFVPNRFVPARQAFVGALATAFCLETARSLFTWYMGNFDGYRS orf144-1 IYGAFAAVPFFLLWLNLLWTLVLGGAVLTSSLSYWQGEAFRRXFDSRGRFDDVLKILLLL 55 orfl44a.pep IYGAFAAVPFFLLWLNLLWTLVLGGAVLTSSLSYWQGEAFRRGFDSRGRFDDVLKILLLL orf144-1 DAAQKEGXALPVQEFRRHINMGYDELGELLEKLARHGYIYSGRQGWVLKTGADSIELNEL orf144a.pep 11:11:11 60 DAAQKEGKALPVQEFRRHINMGYDELGELLEKLARHGYIYSGRQGWVLKTGADSIELNEL orf144-1 FKLFVYRPLPVERDHVNQAVDAVMMPCLQTLNMTLAEFDAQAKKQQQS orf144a.pep FKLFVYRPLPVERDHVNQAVDAVMTPCLQTLNMTLAEFDAQAKKRQ 65 orf144-1

## Homology with a predicted ORF from N. gonorrhoeae

ORF144 shows 91.2% identity over a 136aa overlap with a predicted ORF (ORF144ng) from *N.gonorrhoeae*:

```
5
                     MTFLQRLQGLADNKICAFAWFVVRRFDEERVPQXAASMTFTTLLALVPVLTVMVAVASIF
         orfl44.pep
                                                                              60
                     18181 - 18 181718881881:111:888181 48181818181118818818181118
         orfl44ng
                     MTFLQCWQGSADNKICAFAWFVIRRFSEERVPQAAASMTFTTLLALVPVLTVMVAVASIF
                                                                              60
         orf144.pep
                     PVFDRWSDSFVSFVNQTIVPXGADMVFDYINAFREQANRLTAIGSVMLVVTSLMLIRTID
10
                     orfl44ng
                     PVFDRWSDSFVSFVNQTIVPQGADMVFDYIDAFRDQANRLTAIGSVMLVVTSLMLIRTID
                     NTFNRIWRVXXQRPWM
         orf144.pep
                                                                             136
                     1:11111111111111
15
         orfl44ng
                     NAFNRIWRVNTQRPWMMQFLVYWALLTFGPLSLGVGISFMVGSVQDSVLSSGAQQWADAL
```

The complete length ORF144ng nucleotide sequence <SEQ ID 625> is predicted to encode a protein having amino acid sequence <SEQ ID 626>:

```
MTFLQCWQGS ADNKICAFAW FVIRRFSEER VPQAAASMTF TTLLALVPVL
                     TVMVAVASIF PVFDRWSDSF VSFVNQTIVP QGADMVFDYI DAFRDQANRL
                51
20
                     TAIGSVMLVV TSLMLIRTID NAFNRIWRVN TQRPWMMQFL VYWALLTFGP
                    LSLGVGISFM VGSVQDSVLS SGAQQWADAL KTAARLAFMT LLLWGLYRFV
               151
                    PNREVPARQA FVGALITAFC LETARFLETW YMGNFDGYRS IYGAFAAVPE
                     FLLWLNLLWT LVLGGAVLTS SLSYWQGEAF RRGFDSRGRF DDVLKILLLL
               251
                     DAAQKEGRTL SVQEFRRHIN MGYDELGELL EKLARYGYIY SGRQGWVLKT
               301
25
                     GADSIELSEL FKLFVYRPLP VERDHVNQAV DAVMTPCLQT LNMTLAEFDA
               351
               401
                    QAKKQQQS*
```

Further work revealed the following gonococcal DNA sequence <SEQ ID 627>:

```
1
                   ATGACCTTTT TACAACGTTG GCAAGGTTTG GCGGACAATA AAATCTGTGC
                   ATTTGCATGG TTCGTCATCC GCCGTTTCAG TGAAGAGCGC GTACCGCAGG
30.
               101
                   CAGCGGCGAG CATGACGTTT ACGACACTGC TGGCACTCGT CCCCGTACTG
                   ACCGTAATGG TCGCGGTCGC TTCGATTTTC CCCGTGTTCG ACCGCTGGTC
               151
               201
                   GGATTCGTTC GTCTCCTTCG TCAACCAAAC CATTGTGCCG CAGGGCGCGG
                   ATATGGTGTT CGACTATATC GACGCATTCC GCGATCAGGC AAACCGGCTG
               251
               301
                   ACCGCCATCG GCAGCGTGAT GCTGGTCGTA ACCTCGCTGA TGCTGATTCG
35
               351
                   GACGATAGAC AATGCGTTCA ACCGCATCTG GCGGGTTAAC ACGCAACGCC
               401
                   CCTGGATGAT GCAGTTCCTC GTTTATTGGG CGTTGCTGAC TTTCGGGCCT
               451
                   CGTACTCTCC TCCGGAGCGC AACAATGGGC GGACGCGTTG AAGACGGCGG
               501
               551
                   CAAGGCTGGC TTTCATGACG CTTTTGCTGT GGGGGCTGTA CCGCTTCGTG
40
               601
                   CCCAACCGCT TCGTGCCCGC CCGGCAGGCG TTTGTCGGAG CTTTGATTAC
                   GGCATTCTGC CTGGAGACGG CACGTTTCCT GTTCACCTGG TATATGGGCA
               651
                   ATTTCGACGG CTACCGCTCG ATTTACGGCG CATTTGCCGC CGTGCCGTTT
               701
                   TTCCTGCTGT GGTTAAACCT GCTGTGGACG CTGGTCTTGG GCGGGGCGGT
               751
               801
                   GCTGACTTCG TCGCTGTCTT ATTGGCAGGG CGAGGCCTTC CGCAGGGGAT
45
                   TCGACTCGCG CGGACGGTTT GACGACGTGT TGAAAATCCT GCTGCTTCTG
               851
               901
                   GATGCGGCGC AAAAAGAAGG CCGAACCCTG TCCGTTCAGG AGTTCAGACG
                   GCATATCAAT ATGGGTTACG ATGAATTGGG CGAGCTTTTG GAAAAGCTGG
               951
              1001
                   CGCGGTACGG CTATATCTAT TCCGGCAGAC AGGGCTGGGT TTTGAAAACG
                   GGGGCGGATT CGATTGAGTT GAGCGAACTC TTCAAGCTCT TCGTGTACCG
              1051
50
                   CCCGTTGCct gtggaAAGGG ATCATGTGAA CCAAGCTGtc gaTGCGGTAA
              1101
                   TGACGCCGTG TTTGCAGACT TTGAACATGA CGCTGGCGGA GTTTGACGCT
              1151
                   CAGgcgAAAA AACAGCAGCA GTCTTGA
```

This encodes a variant of ORF144ng, having the amino acid sequence <SEQ ID 628; ORF144ng-1>:

	1	$\mathtt{MTFLQRWQG}L$	ADNKICAFAW	FVIRRFSEER	VPQAAASMTF	TTLLALVPVL
55	51	TVMVAVASIF	PVFDRWSDSF	VSFVNQTIVP	QGADMVFDYI	DAFRDOANRL
	101	TAIGSVMLVV	TSLMLIRTID	NAFNRIWRVN	TORPWMMQFL	VYWALLTFGP
	151	LSLGVGISFM	VGSVQDSVLS	SGAQQWADAL	KTAARLAFMT	LLLWGLYRFV
						IYGAFAAVPF
	251	FLLWLNLLWT	LVLGGAVLTS	SLSYWQGEAF	RRGFDSRGRF	DDVLKILLLL
60	301	DAAQKEGRTL	SVQEFRRHIN	MGYDELGELL	EKLARYGYIY	SGROGWVLKT

351 GADSIELSEL FKLFVYRPLP VERDHVNQAV DAVMTPCLQT LNMTLAEFDA

401 QAKKQQQS\*

ORF144ng-1 and ORF144-1 show 94.1% identity in 406 aa overlap:

```
orf144ng-1.pep MTFLQRWQGLADNKICAFAWFVIRRFSEERVPQAAASMTFTTLLALVPVLTVMVAVASIF
                  5
                  MTFLQRLQGLADNKICAFAWFVVRRFDEERVPQAAASMTFTTLLALVPVLTVMVAVASIF
       orf144-1
       orf144ng-1.pep PVFDRWSDSFVSFVNQTIVPQGADMVFDYIDAFRDQANRLTAIGSVMLVVTSLMLIRTID
                  PVFDRWSDSFVSFVNQTIVPQGADMVFDYINAFREQANRLTAIGSVMLVVTSLMLIRTID
10
       orf144-1
       orfl44ng-1.pep NAFNRIWRVNTQRPWMMQFLVYWALLTFGPLSLGVGISFMVGSVQDSVLSSGAQQWADAL
                  NTFNRIWRVNSQRPWMMQFLVYWALLTFGPLSLGVGISFMVGSVQDAALASGAPQWSGAL
       orf144-1
15
       orf144ng-1.pep KTAARLAFMTLLLWGLYRFVPNRFVPARQAFVGALITAFCLETARFLFTWYMGNFDGYRS
                  RTAATLTFMTLLLWGLYRFVPNRFVPARQAFVGALATAFCLETARSLFTWYMGNFDGYRS
       orf144-1
       orf144ng-1.pep IYGAFAAVPFFLLWLNLLWTLVLGGAVLTSSLSYWQGEAFRRGFDSRGRFDDVLKILLLL
20
                   IYGAFAAVPFFLLWLNLLWTLVLGGAVLTSSLSYWQGEAFRRGFDSRGRFDDVLKILLLL
       orf144-1
        orf144ng-1.pep DAAQKEGRTLSVQEFRRHINMGYDELGELLEKLARYGYIYSGRQGWVLKTGADSIELSEL
                   25
                  DAAQKEGKALPVQEFRRHINMGYDELGELLEKLARHGYIYSGRQGWVLKTGADSIELNEL
        orf144-1
        orf144ng-1.pep FKLFVYRPLPVERDHVNQAVDAVMTPCLQTLNMTLAEFDAQAKKQQQS
                   FKLFVYRPLPVERDHVNQAVDAVMTPCLQTLNMTLAEFDAQAKKRQ
30
        orf144-1
```

On this basis of this analysis, including the identification of several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### 35 Example 75

40

The following partial DNA sequence was identified in N. meningitidis <SEQ ID 629>:

```
1 ..AGACACGCCC GCCGCATCCG CATCGACACC GCCATCAACC CCGAACTGGA
51 AGCCCTCGCC GAACACCTCC ACTACCAATG GCAGGGCTTC CTCTGGCTCA
101 GCACCGATAT GCGTCAGGAA ATTTCCGCCC TCGTCATCCT GCTGCAACGC
151 ACCCGCCGCA AATGGCTGGA TGCCCACGAA CGCCAACACC TGCGCCAAAG
201 CCTGCTTGAA ACACGGGAAC ACGGCTGA
```

This corresponds to the amino acid sequence <SEQ ID 630; ORF146>:

- 1 ..RHARRIRIDT AINPELEALA EHLHYQWQGF LWLSTDMRQE ISALVILLQR 51 TRRKWLDAHE RQHLRQSLLE TREHG\*
- 45 Further work revealed the complete nucleotide sequence <SEQ ID 631>:

```
ATGAACACCT CGCAACGCAA CCGCCTCGTC AGCCGCTGGC TCAACTCCTA
                    CGAACGCTAC CGCTACCGCC GCCTCATCCA CGCCGTCCGG CTCGGCGGGG
                51
                    CCGTCCTGTT CGCCACCGCC TCCGCCCGGC TGCTCCACCT CCAACACGGC
               101
                    GAGTGGATAG GGATGACCGT CTTCGTCGTC CTCGGCATGC TCCAGTTTCA
               151
                    AGGGGCGATT TACTCCAAGG CGGTGGAACG TATGCTCGGC ACGGTCATCG
50
               201
                    GGCTGGGCGC GGGTTTGGGC GTTTTATGGC TGAACCAGCA TTATTTCCAC
               251
                    GGCAACCTCC TCTTCTACCT CACCGTCGGC ACGGCAAGCG CACTGGCCGG
               301
                    CTGGGCGGCG GTCGGCAAAA ACGGCTACGT CCCTATGCTG GCAGGGCTGA
               351
                    CGATGTGTAT GCTCATCGGC GACAACGGCA GCGAATGGCT CGACAGCGGA
               401
               451 CTCATGCGCG CCATGAACGT CCTCATCGGC GCGGCCATCG CCATCGCCGC
55
```

	501 551	CGCCAAACTG CCGACAACCT	CTGCCGCTGA GGCCGACTGC	AATCCACACT AGCAAAATGA	GATGTGGCGT TTGCCGAAAT	TTCATGCTTG
	601	AGGCGCATGA	CCCGCGAACG	CCTCGAGGAG	AACATGGCGA	AAATGCGCCA
_	651	AATCAACGCA	CGCATGGTCA	AAAGCCGCAG	CCATCTCGCC	GCCACATCGG
<b>.</b>	701	GCGAAAGCCG	CATCAGCCCC	GCCATGATGG	AAGCCATGCA	GCACGCCCAC
	751	CGTAAAATCG	TCAACACCAC	CGAGCTGCTC	CTGACCACCG	CCGCCAAGCT
	801	GCAATCTCCC	AAACTCAACG	GCAGCGAAAT	CCGGCTGCTT	GACCGCCACT
	851	TCACACTGCT	CCAAACCGAC	CTGCAACAAA	CCGTCGCCCT	TATCAACGCC
10	901	AGACACGCCC	GCCGCATCCG	CATCGACACC	GCCATCAACC	CCGAACTGGA
10	951	AGCCCTCGCC	GAACACCTCC	ACTACCAATG	GCAGGGCTTC	CTCTGGCTCA
	1001	GCACCAATAT	GCGTCAGGAA	ATTTCCGCCC	TCGTCATCCT	GCTGCAACGC
]	1051	ACCCGCCGCA	AATGGCTGGA	TGCCCACGAA	CGCCAACACC	TGCGCCAAAG
:			ACACGGGAAC			DAMAGGGG

# This corresponds to the amino acid sequence <SEQ ID 632; ORF146-1>:

```
1 MNTSQRNRLV SRWLNSYERY RYRRLIHAVR LGGAVLFATA SARLLHLQHG
51 EWIGMTVFVV LGMLQFQGAI YSKAVERMLG TVIGLGAGLG VLWLNQHYFH
101 GNLLFYLTVG TASALAGWAA VGKNGYVPML AGLTMCMLIG DNGSEWLDSG
151 LMRAMNVLIG AAIAIAAAKL LPLKSTLMWR FMLADNLADC SKMIAEISNG
201 RRMTRERLEE NMAKMRQINA RWVKSRSHLA ATSGESRISP AMMEAMQHAH
251 RKIVNTTELL LTTAAKLQSP KLNGSEIRLL DRHFTLLQTD LQQTVALING
301 RHARRIRIDT AINPELEALA EHLHYQWQGF LWLSTNMRQE ISALVILLQR
351 TRKWLDAHE RQHLRQSLLE TREHG*
```

Computer analysis of this amino acid sequence gave the following results:

## Homology with a predicted ORF from N.meningitidis (strain A)

ORF146 shows 98.6% identity over a 74aa overlap with an ORF (ORF146a) from strain A of N. meningitidis:

	orf146.pep			RHAI	10 RRIRIDTAINP	20 ELEALAEHLI	30 HYOWOGE		
30	orf146a	KLNGSEIRLLDRHE	TLLQTDLQQ	1111		LITTLE	HĪHĪH		
		280	290	300	310	320	330		
		40	50	60	. 70				
0.5	orf146.pep	LWLSTDMRQEISAL	VILLORTRR	KWLDAHERQH1	LROSLLETREH	GX			
35	orf146a	{							
		340	350	360	370				

## The complete length ORF146a nucleotide sequence <SEQ ID 633> is:

	1	ATGAACACCT	CGCAACGCAA	CCGCCTCGTC	AGCCGCTGGC	TCAACTCCTA
40	51	CGAACGCTAC	CGCTACCGCC	GCCTCATCCA	CGCCGTCCGG	CTCGGCGGGG
	101	CCGTCCTGTT	CGCCACCGCC	TCCGCCCGGC	TGCTCCACCT	CCAACACGGC
	151	GAGTGGATAG	GGATGACCGT		CTCGGCATGC	
	201	AGGGGCGATT	TACTCCAAGG	CGGTGGAACG	TATGCTCGGC	ACGGTCATCG
	251	GGCTGGGCGC	GGGTTTGGGC	GTTTTATGGC	TGAACCAGCA	TTATTTCCAC
45	301	GGCAACCTCC	TCTTCTACCT	CACCGTCGGC	ACGGCAAGCG	CACTGGCCGG
	351	CTGGGCGGCG	GTCGGCAAAA	ACGGCTACGT	CCCTATGCTG	GCGGGGCTGA
	401	CGATGTGCAT	GCTCATCGGC	GACAACGGCA	GCGAATGGTT	CCACACCCCC
	451	CTGATGCGCG	CGATGAACGT	CCTCATCGGC	GCGGCCATCG	CCATCGCCGC
	501	CGCCAAACTG	CTGCCGCTGA	AATCCACACT	GATGTGGCGT	TTCATCCTTC
50	551	CCGACAACCT	GACCGACTGC	AGCAAAATGA	TTGCCGAAAT	CACCAACCCC
	601	AGGCGCATGA	CCCGCGAACG	CCTCGAAGAG	AACATGGCGA	AAATCCCCCA
	651	AATCAACGCA	CGCATGGTCA	AAAGCCGCAG	CCACCTCGCC	CCCACATCCC
	701	GCGAAAGCCG	CATCAGCCCC	GCCATGATGG	AAGCCATGCA	CCACCATCGG
	751	CGTAAAATTG	TCAACACCAC	CGAGCTGCTC	CTGACCACCG	CCCCCAACCO
55	801	GCAATCTCCC	AAACTCAACG	GCAGCGAAAT	CCGCCTCCTT	GACCGCCACT
	851	TCACACTGCT	CCAAACCGAC	CTCCDACAAA	CCGTCGCCCT	TATCAACGGC
	901	AGACACGCCC	GCCGCATCCG	CATCCACACA	GCCATCAACC	CCCARCOGC
	951	AGCCCTCGCC	GAACACCTCC	ACTACCAATC	GCAGGGCTTC	CUGAACTGGA
	1001	GCACCAATAT	GCGTCAGGAA	ACTACCAATG	TCGTCATCCT	CTCTGGCTCA
60	1051	ACCCGCCGCA	AATGGCTGGA	TCCCCACCAA	CGCCAACACC	GCTGCAACGC
<del>*</del> -	1101	CCTGCTTGAA	ACACGGGAAC	ACACEMCA ACACEMCA	CGCCAACACC	TGCGCCAAAG
	~~~	00100110121	COGGHAC	ACAGITGA.		

This encodes a protein having amino acid sequence <SEQ ID 634>:

```
MNTSQRNRLV SRWLNSYERY RYRRLIHAVR LGGAVLFATA SARLLHLQHG
51 EWIGMTVFVV LGMLQFQGAI YSKAVERMLG TVIGLGAGLG VLWLNQHYFH
101 GNLLFYLTVG TASALAGWAA VGKNGYVPML AGLTMCMLIG DNGSEWFDSG
5 151 LMRAMNVLIG AAIAIAAAKL LPLKSTLMWR FMLADNLTDC SKMIAEISNG
201 RRMTRERLEE NMAKMRQINA RMVKSRSHLA ATSGESRISP AMMEAMQHAH
251 RKIVNTTELL LTTAAKLQSP KLNGSEIRLL DRHFTLLQTD LQQTVALING
301 RHARRIRIDT AINPELEALA EHLHYQWQGF LWLSTNMRQE ISALVILLQR
351 TRRKWLDAHE RQHLRQSLLE TREHS*
```

ORF146a and ORF146-1 show 99.5% identity in 374 aa overlap:

```
MNTSQRNRLVSRWLNSYERYRYRRLIHAVRLGGAVLFATASARLLHLQHGEWIGMTVFVV
        orf146a.pep
                   orf146-1
                   MNTSQRNRLVSRWLNSYERYRYRRLIHAVRLGGAVLFATASARLLHLQHGEWIGMTVFVV
15
                   {\tt LGMLQFQGAIYSKAVERMLGTVIGLGAGLGVLWLNQHYFHGNLLFYLTVGTASALAGWAA}
        orf146a.pep
                    orf146-1
                   LGMLQFQGAIYSKAVERMLGTVIGLGAGLGVLWLNQHYFHGNLLFYLTVGTASALAGWAA
                   VGKNGYVPMLAGLTMCMLIGDNGSEWFDSGLMRAMNVLIGAAIAIAAAKLLPLKSTLMWR
        orf146a.pep
                    20
        orf146-1
                   VGKNGYVPMLAGLTMCMLIGDNGSEWLDSGLMRAMNVLIGAAIAIAAAKLLPLKSTLMWR
        orf146a.pep
                   FMLADNLTDCSKMIAEISNGRRMTRERLEENMAKMRQINARMVKSRSHLAATSGESRISP
                    25
        orf146-1
                   FMLADNLADCSKMIAEISNGRRMTRERLEENMAKMRQINARMVKSRSHLAATSGESRISP
        orf146a.pep
                   {\tt AMMEAMQHAHRKIVNTTELLLTTAAKLQSPKLNGSEIRLLDRHFTLLQTDLQQTVALING}
                    orf146-1
                   {\tt AMMEAMQRAHRKIVNTTELLLTTAAKLQSPKLNGSEIRLLDRHFTLLQTDLQQTVALING}
30
                   RHARRIRIDTAIN \texttt{PELEALAEHLHYQWQGFLWLSTNMRQEISALVILLQRTRRKWLDAHE}
        orf146a.pep
                    orf146-1
                   RHARRIRIDTAINPELEALAEHLHYQWQGFLWLSTNMRQEISALVILLQRTRRKWLDAHE
35
                   RQHLRQSLLETREHSX
        orf146a.pep
                    1111111111111111
                   ROHLROSLLETREHGX
        orf146-1
```

#### Homology with a predicted ORF from N.gonorrhoeae

40 ORF146 shows 97.3% identity over a 75aa overlap with a predicted ORF (ORF146ng) from N.gonorrhoeae:

An ORF146ng nucleotide sequence <SEQ ID 635> was predicted to encode a protein having amino

50 acid sequence <SEQ ID 636>:

```
1 MSGVRFPSPA PIPSTDPPSG SLCFFTFPLQ TASDMNSSQR KRLSGRWLNS
                    YERYRHRRLI HAVRLGGTVL FATALARLLH LOHGEWIGMT VFVVLGMLQF
                51
                    QGAIYSNAVE RMLGTVIGLG AGLGVLWLNQ HYFHGNLLFY LTIGTASALA
               101
                    GWAAVGKNGY VPMLAGLTMC MLIGDNGSEW LDSGLMRAMN VLIGAAIAIA
               151
55
                    AAKLLPLKST LMWRFMLADN LADCSKMIAE ISNGRRMTRE RLEQNMVKMR
               201
                    QINARMVKSR SHLAATSGES RISPSMMEAM QHAHRKIVNT TELLLTTAAK
               251
                   LQSPKLNGSE IRLLDRHFTL LQTDLQQTAA LINGRHARRI RIDTAINPEL
               301
                    EALAEHLHYQ WQGFLWLSTN MRQEISALVI PLQRTRRKWL DAHERQHLRQ
               351
                    SLLETREHG*
                401
```

60 Further work revealed the following gonococcal DNA sequence <SEQ ID 637>:

```
ATGAACTCCT CGCAACGCAA ACGCCTTTCC GgccGCTGGC TCAACTCCTA
               51
                   CGAACGCTac cGCCaccGCC GCCTCATACA TGCCGTGCGG CTCGGCggaa
                   ccgtCCTGTT CGCCACCGCA CTCGCCCGgc tACTCCACCT CCAacacggc
              101
              151
                   GAATGGATAG GGAtgaCCGT CTTCGTCGTC CTCGGCATGC TCCAGTTCCA
 5
              201
                   AGGCgcgatt tActccaacg cggtgGAacg taTGctcggt acggtcatcg
              251
                   ggctgGGCGC GGGTTTGGgc gTTTTATGGC TGAACCAGCA TTAtttccac
                   ggcaacCTcc tettetacet gaccategge aeggeaageg caetggeegg
              301
                   ctGGGCGGCG GTCGGCAAAA acggctacgt ccctatgctg GCGGGGctgA
              351
              401
                   CGATGTGCAT getcatcggc gACAACGGCA GCGAATGGCT CGACAGCGGC
10
              451
                   CTGATGCGCG CGATGAACGT CCTCATCGGC GCCGCCATCG CCATTGCCGC
                   CGCCAAACTG CTGCCGCTGA AATCCACACT GATGTGGCGT TTCATGCTTG
              501
                   CCGACAACCT GGCCGACTGC AGCAAAATGA TTGCCGAAAT CAGCAACGGC
              551
              601
                   AGGCGTATGA CGCGCGAACG TTTGGAGCAG AATATGGTCA AAATGCGCCA
                   AATCAACGCA CGCATGGTCA AAAGCCGCAG CCACCTCGCC GCCACATCGG
              651
15
              701
                  GCGAAAGCCG CATCAGCCCC TCCATGATGG AAGCCATGCA GCACGCCCAC
                   CGCAAAATCG TCAACACCAC CGAGCTGCTC CTGACCACCG CCGCCAAGCT
              751
                  GCAATCTCCC AAACTCAACG GCAGCGAAAT CCGGCTGCTC GACCGCCACT
              801
              851
                  TCACACTGCT CCAAACCGAC CTGCAACAAA CCGCCGCCCT CATCAACGGC
                  AGACACGCCC GCCGCATCCG CATCGACACC GCCATCAACC CCGAACTGGA
AGCCCTCGCC GAACACCTCC ACTACCAATG GCAGGGCTTC CTCTGGCTCA
              901
20.
              951
                  GCACCAATAT GCGTCAGGAA ATTTCCGCCC TCGTCATCCT GCTGCAACGC
             1001
             1051
                  ACCCGCCGCA AATGGCTGGA TGCCCACGAA CGCCAACACC TGCGCCAAAG
             1101
                  CCTGCTTGAA ACACGGGAAC ACGGCTGA
     This corresponds to the amino acid sequence <SEQ ID 638; ORF146ng-1>:
25
                  MNSSQRKRLS GRWLNSYERY RHRRLIHAVR LGGTVLFATA LARLLHLQHG
               51
                  EWIGMTVFVV LGMLQFQGAI YSNAVERMLG TVIGLGAGLG VLWLNQHYFH
                  GNLLFYLTIG TASALAGWAA VGKNGYVPML AGLTMCMLIG DNGSEWLDSG
              101
              151
                  LMRAMNVLIG AAIAIAAAKL LPLKSTLMWR FMLADNLADC SKMIAEISNG
                  RRMTRERLEQ NMVKMRQINA RMVKSRSHLA ATSGESRISP SMMEAMQHAH
              201
30
                  RKIVNTTELL LTTAAKLQSP KLNGSEIRLL DRHFTLLQTD LQQTAALING
              251
                  RHARRIRIDT AINPELEALA EHLHYQWQGF LWLSTNMRQE ISALVILLQR
                  TRRKWLDAHE ROHLROSLLE TREHG*
     ORF146ng-1 and ORF146-1 show 96.5% identity in 375 aa overlap
                       MNTSQRNRLVSRWLNSYERYRYRRLIHAVRLGGAVLFATASARLLHLQHGEWIGMTVFVV
         orf146-1.pep
35 -
                       orfl46ng-1
                       MNSSQRKRLSGRWLNSYERYRHRRLIHAVRLGGTVLFATALARLLHLQHGEWIGMTVFVV
         orf146-1.pep
                       LGMLQFQGAIYSKAVERMLGTVIGLGAGLGVLWLNQHYFHGNLLFYLTVGTASALAGWAA
                       40
         orf146ng-1
                       LGMLQFQGAIYSNAVERMLGTVIGLGAGLGVLWLNQHYFHGNLLFYLTIGTASALAGWAA
         orf146-1.pep
                       VGKNGYVPMLAGLTMCMLIGDNGSEWLDSGLMRAMNVLIGAAIAIAAAKLLPLKSTLMWR
                       orf146ng-1
                       VGKNGYVPMLAGLTMCMLIGDNGSEWLDSGLMRAMNVLIGAAIAIAAAKLLPLKSTLMWR
45
                       {\tt FMLADNLADCSKMIAEISNGRRMTRERLEENMAKMRQINARMVKSRSHLAATSGESRISP}
         orf146-1.pep
                       orf146ng-1
                       FMLADNLADCSKMIAEISNGRRMTRERLEQNMVKMRQINARMVKSRSHLAATSGESRISP
50
         orf146-1.pep
                       AMMEAMQHAHRKIVNTTELLLTTAAKLQSPKLNGSEIRLLDRHFTLLQTDLQQTVALING
                       SMMEAMQHAHRKIVNTTELLLTTAAKLQSPKLNGSEIRLLDRHFTLLQTDLQQTAALING
         orf146ng-1
         orf146-1.pep
                       RHARRIRIDTAINPELEALAEHLHYQWQGFLWLSTNMRQEISALVILLQRTRRKWLDAHE
55
                       orf146ng-1
                       RHARRIRIDTAINPELEALAEHLHYQWQGFLWLSTNMRQEISALVILLQRTRRKWLDAHE
         orf146-1.pep
                       RQHLRQSLLETREHGX
                       60
         orf146ng-1
                       RQHLRQSLLETREHGX
```

Furthermore, ORF146ng-1 shows homology with a hypothetical E.coli protein:

sp|P33011|YEEA\_ECOLI HYPOTHETICAL 40.0 KD PROTEIN IN COBU-SBMC INTERGENIC REGION >gi|1736674|gnl|PID|d1016553 (D90838) ORF\_ID:0348#20; similar to [SwissProt Accession Number P33011] [Escherichia coli] >gi|1736682|gnl|PID|d1016560 (D90839) ORF\_ID:0348#20; similar to [SwissProt Accession Number P33011] [Escherichia coli]

```
>gi|1788318 (AE000292) f352; 100% identical to fragment YEEA ECOLI SW: P33011 but
          has 203 additional C-terminal residues [Escherichia coli] Length = 352
           Score = 109 bits (271), Expect = 2e-23
           Identities = 89/347 (25%), Positives = 150/347 (42%), Gaps = 21/347 (6%)
5
          Query: 20 YRHRRLIHAVRLGGTVLFATALARLIHLQHGEWIGMTVFVVLGMLQFQGAIYSNAVERML 79
                     YRH R++H R+ L + RL + W +T+ V++G + F G + A ER+
          Sbjct: 15 YRHYRIVHGTRVALAFLLTFLIIRLFTIPESTWPLVTMVVIMGPISFWGNVVPRAFERIG 74
10
          Query: 80 GTVIGLGAGLGVLWLNQHYFHGNLLFYLTIGTASALAGWAAVGKNGYVPMLAGLTMCMLI 139
                     GTV+G GL L L + A L GW A+GK Y +L G+T+ +++
          Sbjct: 75 GTVLGSILGLIALQLE---LISLPLMLVWCAAAMFLCGWLALGKKPYQGLLIGVTLAIVV 131
          Query: 140 GDNGSEWLDSGLMRAMNVLIGXXXXXXXXKLLPLKSTLMWRFMLADNLADCSKMIAEISN 199
15
                         E + D + L R + + V + + G + P + + + WR LA + L + + + +
          Sbjct: 132 GSPTGE-IDTALWRSGDVILGSLLAMLFTGIWPQRAFIHWRIQLAKSLTEYNRVYQSAFS 190
          Query: 200 GRRMTRERLEQNMVKMRQINARMVKSRSHLAATSGESRISPSMMEAMQHAHRKIVNXXXX 259
                        + R RLE ++ K+
                                          VK R +A S E+RI S+ E +Q +R +V
20
          Sbjct: 191 PNLLERPRLESHLQKLL---TDAVKMRGLIAPASKETRIPKSIYEGIQTINRNLVCMLEL 247
          Query: 260 XXXXXXXQSPK---LNGSEIRLLDRHFXXXXXXXXXAALINGRHARRIRIDTAINPEL 316
                                   LN ++R D
                                                          AL G
          Sbjct: 248 QINAYWATRPSHFVLLNAQKLR--DTQHMMQQILLSLVHALYEGNPQPVFANTEKLNDAV 305
25
          Query: 317 EALAEHL--HYQWQ-----GFLWLSTNMRQEISALVILLQRTRRK 354
                     E L + L H+ +
                                         G++WL+
                                                   ++ L L+ R RK
          Sbjct: 306 EELRQLLNNHHDLKVVETPIYGYVWLNMETAHQLELLSNLICRALRK 352
```

On the basis of this analysis, including the identification of several transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 76

30

The following partial DNA sequence was identified in N.meningitidis <SEQ ID 639>

```
..GCCGAAGACA CGCGCGTTAC CGCACAGCTT TTGAGCGCGT ACGGCATTCA
35
                  51
                        GGGCAAACTC GTCAGTGTGC GCGAACACAA CGAACGGCAG ATGGCGGACA
                        AGATTGTCGG CTATCTTTCA GACGGCATGG TTGTGGCACA GGTTTCCGAT
                 101
                 151
                        GCGGGTACGC CGGCCGTGTG CGACCCGGGC GCGAAACTCG CCCGCCGCGT
                        GCGTGAGGCC GGGTTTAAAG TCGTTCCCGT CGTGGGCGCA AC.GCGGTGA
TGGCGGCTTT GAGCGTGGCC GGTGTGGAAG GATCCGATTT TTATTTCAAC
                 201
                 251
40
                        GGTTTTGTAC CGCCGAAATC GGGAGAACGC AGGAAACTGT TTGCCAAATG
                 301
                 351
                        GGTGCGGCCG GCGTTTCCTA TCGTCATGTT TGAAACGCCG CACCGCATCG
                        GTGCAGCGCT TGCCGATATG GCGGAACTGT TCCCCGAACG CCGATTAATG
                 401
                 451
                        CTGGCGCGC AAATTACGAA AACGTTTGAA ACGTTCTTAA GCGGCACGGT
                 501
                        TGGGGAAATT CAGACGGCAT TGTCTGCCGA CGGCGACCAA TCGCGCGGCG
45
                        AGATGGTGTT GGTGCTTTAT CCGGCGCAGG ATGAAAAACA CGAAGGCTTG
                 551
                 601
                        TCCGAGTCCG CGCAAAACAT CATGAAAATC CTCACAGCCG AGCTGCCGAC
                        CAAACAGGCG GCGGAGCTTG CTGCCAAAAT CACGGGCGAG GGAAAGAAAG
                 651
                 701
                        CTTTGTACGA T..
```

This corresponds to the amino acid sequence <SEQ ID 640; ORF147>:

```
1 ..AEDTRVTAQL LSAYGIQGKL VSVREHNERQ MADKIVGYLS DGMVVAQVSD
51 AGTPAVCDPG AKLARRVREA GFKVVPVVGA XAVMAALSVA GVEGSDFYFN
101 GFVPPKSGER RKLFAKWVRA AFPIVMFETP HRIGAALADM AELFPERRLM
151 LAREITKTFE TFLSGTVGEI QTALSADGDQ SRGEMVLVLY PAQDEKHEGL
201 SESAQNIMKI LTAELPTKQA AELAAKITGE GKKALYD..
```

55 Further work revealed the complete nucleotide sequence <SEQ ID 641>:

- 1 ATGTTTCAGA AACATTTGCA GAAAGCCTCC GACAGCGTCG TCGGAGGGAC
  51 ATTATACGTG GTTGCCACGC CCATCGGCAA TTTGGCGGAC ATTACCCTGC
- 101 GCGCTTTGGC GGTATTGCAA AAGGCGGACA TCATCTGTGC CGAAGACACG 151 CGCGTTACCG CACAGCTTTT GAGCGCGTAC GGCATTCAGG GCAAACTCGT

	201 CA	GTGTGCGC GAACACAACG AACGGCAGAT GGCGGACAAG ATTGTCGGCT
		CTTTCAGA CGGCATGGTT GTGGCACAGG TTTCCGATGC GGGTACGCCG
		CGTGTGCG ACCCGGGCGC GAAACTCGCC CGCCGCGTGC GTGAGGCCGG
5		TTAAAGTC GTTCCCGTCG TGGGCGCAAG CGCGGTGATG GCGGCTTTGA
3		GTGGCCGG TGTGGAAGGA TCCGATTTTT ATTTCAACGG TTTTGTACCG GAAATCGG GAGAACGCAG GAAACTGTTT GCCAAATGGG TGCGGGCGGC
		TTCCTATC GTCATGTTTG AAACGCCGCA CCGCATCGGT GCGGCCGCC
	551 CC	GATATGGC GGAACTGTTC CCCGAACGCC GATTAATGCT GGCGCGCGAA
	601 AT	TACGAAAA CGTTTGAAAC GTTCTTAAGC GGCACGGTTG GGGAAATTCA
10		CGGCATTG TCTGCCGACG GCAACCAATC GCGCGGCGAG ATGGTGTTGG
	701 TG	CTTTATCC GGCGCAGGAT GAAAAACACG AAGGCTTGTC CGAGTCCGCG
	751 CA	AAACATCA TGAAAATCCT CACAGCCGAG CTGCCGACCA AACAGGCGGC
		AGCTTGCT GCCAAAATCA CGGGCGAGGG AAAGAAAGCT TTGTACGATC
	851 <b>T</b> G	GCTCTGTC TTGGAAAAAC AAATAG
15	This corresponds to	the amino acid sequence <seq 642;="" id="" orf147-1="">:</seq>
13	Tims corresponds to	the armie and sequence ODQ ID 042, Old 147-17.
	1 MF	QKHLQKAS DSVVGGTLYV VATPIGNLAD ITLRALAVLQ KADIICAEDT
		TAQLLSAY GIQGKLVSVR EHNERQMADK IVGYLSDGMV VAQVSDAGTP
•	101 AV	CDPGAKLA RRVREAGFKV VPVVGASAVM AALSVAGVEG SDFYFNGFVP
	151 PK	SGERRKLF AKWVRAAFPI VMFETPHRIG ATLADMAELF PERRLMLARE
20	201 IT	KTFETFLS GTVGEIQTAL SADGNQSRGE MVLVLYPAQD EKHEGLSESA
	251 QN	IMKILTAE LPTKQAAELA AKITGEGKKA LYDLALSWKN K*
	(1	of this amine said seemen as an Al- C-11in a said
	Computer analysis	of this amino acid sequence gave the following results:
	Homology with hy	pothetical protein ORF286 of E.coli (accession number U18997)
	ODE147 - 1 F - 1	ODFOOC 1 1 200/ - 11 // 1 000 1
	ORF147 and E.coli	ORF286 protein show 36% aa identity in 237aa overlap:
25		
25	Orf147: 1	AEDTRVTAQLLSAYGIQGKLVSVREHNERQMADKIVGYLSDGMVVAQVSDAGTPAVCDPG 60
	Orf286. 43	AEDTR T LL +GI +L ++ +HNE+Q A+ ++ L +G +A VSDAGTP + DPG AEDTRHTGLLLQHFGINARLFALHDHNEQQKAETLLAKLQEGQNIALVSDAGTPLINDPG 102
	Q11280. <b>4</b> 5	PEDITULI GENERALI GENERALI PEDUNDO Ó VERT PEDEVE DO GALLE LA DAGI LA TRADA DAGI LA TRADA DE COMO DE CO
	Orf147: 61	AKLARRVREXXXXXXXXXXXXXXXXXXXXXXXXEGSDFYFNGFVPPKSGERRKLFAKWVRA 120
30	·	L R RE F + GF+P KS RR
	Orf286: 103	YHLVRTCREAGIRVVPLPGPCAAITALSAAGLPSDRFCYEGFLPAKSKGRRDALKAIEAE 162
	0.5140101	
	Orf147: 121	AFPIVMFETPHRIGAALADMAELFPERR-LMLAREITKTFETFLSGTVGEIQTALSADGD 179
35	Orf286 163	++ +E+ HR+ +L D+ + E R ++LARE+TKT+ET VGE+ + D + PRTLIFYESTHRLLDSLEDIVAVLGESRYVVLARELTKTWETIHGAPVGELLAWVKEDEN 222
55	011200. 103	TRIBITIESTERNOODSUBDIVAVEGESRIVVUAREDIRIWEIINGAFVGEDLAWVREDEN 222
	Orf147: 180	QSRGEMVLVLYPAQDEKHEGLSESAQNIMKILTAELPTKQAAELAAKITGEGKKALY 236
		+ +GEMVL++ + E L A + +L AELP K+AA LAA+I G K ALY
40	Orf286: 223	RRKGEMVLIV-EGHKAQEEDLPADALRTLALLQAELPLKKAAALAAEIHGVKKNALY 278
40		
	Homology with a p	oredicted ORF from N.meningitidis (strain A)
	ODE147 -1 06	CO/ 11-414 027 1 14 ODDES C
	OKF 14 / snows 96.0	6% identity over a 237aa overlap with ORF75a from strain A of N. meningitidis:
	·	
	orf147.pep	10 20 30
45	oril47.pep	
. 73	orf75a	
	022700	20 30 40 50 60 70
	•	
<b></b>		40 50 60 70 80 90
50	orf147.pep	MADKIVGYLSDGMVVAQVSDAGTPAVCDPGAKLARRVREAGFK <u>VVPVVGAXAVMAALSVA</u>
	676 -	
	orf75a	MADKIVGYLSDGMVVAQVSDAGTPAVCDPGAKLARRVREVGFKVVPVVGASAVMAALSVA
		80 90 100 110 120 130
55	•	100 110 120 130 140 150
	orf147.pep	GVEGSDFYFNGFVPPKSGERRKLFAKWVRAAFPIVMFETPHRIGAALADMAELFPERRLM
		11 111111111111111111111111111111111111
	orf75a	GVAGSDFYFNGFVPPKSGERRKLFAKWVRVAFPVVMFETPHRIGATLADMAELFPERRLM
60		140 150 160 170 180 190
OU		160 170 180 190 200 210

160 170 180 190 200 210 LAREITKTFETFLSGTVGEIQTALSADGDQSRGEMVLVLYPAQDEKHEGLSESAQNIMKI

WO 99/24578 PCT/IB98/01665

-357-

	orf75a	 LAREITKT		 EIQTALAADG		 LYPAQDEKHE	EGLSESAQNIMKI
		200	210	220	230	240	250
5		2	20 2	:30			
	orf147.pep	_	QAAELAAKIT				
	45.5	,,,,,,,,	1111111111		N. T. ORTHOLIES	•	
	orf75a		_	GEGKKALYDI			
		260	270	280	290		

ORF147a is identical to ORF75a, which includes aa 56-292 of ORF75.

## Homology with a predicted ORF from N.gonorrhoeae

ORF147 shows 94.1% identity over a 237aa overlap with a predicted ORF (ORF147ng) from N. gonorrhoeae:

15	orf147.pep	AEDTRVTAQLLSAYGIQGKLVSVREHNERQ	30
	orf147ng	TLYVVATPIGNLADITLRALAVLQKADIICAEDTRVTAQLLSAYGIQGRLVSVREHNERQ	85
	orf147.pep	MADKIVGYLSDGMVVAQVSDAGTPAVCDPGAKLARRVREAGFKVVPVVGAXAVMAALSVA	90
20	orf147ng	MADKVIGFLSDGLVVAQVSDAGTPAVCDPGAKLARRVREAGFKVVPVVGASAVMAALSVA	145
	orf147.pep	GVEGSDFYFNGFVPPKSGERRKLFAKWVRAAFPIVMFETPHRIGAALADMAELFPERRLM	150
25	orf147ng	GVAESDFYFNGFVPPKSGERRKLFAKWVRAAFPVVMFETPHRIGATLADMAELFPERRLM	205
23	orf147.pep	LAREITKTFETFLSGTVGEIQTALSADGDQSRGEMVLVLYPAQDEKHEGLSESAQNIMKI	210
	orf147ng	LAREITKTFETFLSGTVGEIQTALAADGNQSRGEMVLVLYPAQDEKHEGLSESAQNAMKI	265
30	orf147.pep	LTAELPTKQAAELAAKITGEGKKALYD 237	
	orf147ng	LAAELPTKQAAELAAKITGEGKKALYDLALSWKNK 300	

An ORF147ng nucleotide sequence <SEQ ID 643> was predicted to encode a protein having amino acid sequence <SEQ ID 644>:

```
35 1 MSVFQTAFFM FQKHLQKASD SVVGGTLYVV ATPIGNLADI TLRALAVLQK
51 ADIICAEDTR VTAQLLSAYG IQGRLVSVRE HNERQMADKV IGFLSDGLVV
101 AQVSDAGTPA VCDPGAKLAR RVREAGFKVV PVVGASAVMA ALSVAGVAES
151 DFYFNGFVPP KSGERRKLFA KWVRAAFPVV MFETPHRIGA TLADMAELFP
201 ERRLMLAREI TKTFETFLSG TVGEIQTALA ADGNQSRGEM VLVLYPAQDE
40 251 KHEGLSESAQ NAMKILAAEL PTKQAAELAA KITGEGKKAL YDLALSWKNK
301 *
```

Further work revealed the following gonococcal DNA sequence <SEQ ID 645>:

	1	ATGTTTCAGA	AACACTTGCA	GAAAGCCTCC	GACAGCGTCG	TCGGAGGGAC
	51	ATTATACGTG	GTTGCCACGC	CCATCGGCAA	TTTGGCAGAC	ATTACCCTGC
45	101	GCGCTTTGGC	GGTATTGCAA	AAGGCGGACA	TCATTTGTGC	CGAAGACACG
	151	CGCGTTACTG	CGCAGCTTTT	GAGCGCGTAC	GGCATTCAGG	GCAGGTTGGT
	201	CAGTGTGCGC	GAACACAACG	AGCGGCAGAT	GGCGGACAAG	GTAATCGGTT
	251	TCCTTTCAGA	CGGCCTGGTT	GTGGCGCAGG	TTTCCGATGC	GGGTACGCCG
	301	GCCGTGTGCG	ACCCGGGCGC	GAAACTCGCC	CGCCGCGTGC	GCGAAGCAGG
50	351	GTTCAAAGTC	GTTCCCGTCG	TGGGCGCAAG	CGCGGTAATG	GCGGCGTTGA
•	401	GTGTGGCCGG	TGTGGCGGAA	TCCGATTTTT	ATTTCAACGG	TTTTGTACCG
	451	CCGAAATCGG	GCGAACGTAG	GAAATTGTTT	GCCAAATGGG	TGCGGGCGGC
	501	ATTTCCTGTC	GTCATGTTTG	AAACGCCGCA	CCGAATCGGG	GCAACGCTTG
	551	CCGATATGGC	GGAATTGTTC	CCCGAACGCC	GTCTGATGCT	GGCGCGCGAA
55	601	ATCACGAAAA	CGTTTGAAAC	GTTCTTAAGC	GGCACGGTTG	GGGAAATTCA
	651	GACGGCATTG	GCGGCGGACG	GCAACCAATC	GCGCGGCGAG	ATGGTGTTGG
	701	TGCTTTATCC	GGCGCAGGAT	GAAAAACACG	AAGGCTTGTC	CGAGTCTGCG
	751				CTGCCGACCA	
	801				CAAAAAGGCT	
60	851		GTGGAAAAAC			

5

This corresponds to the amino acid sequence <SEQ ID 646; ORF147ng-1>:

```
1 MFQKHLQKAS DSVVGGTLYV VATPIGNLAD ITLRALAVLQ KADIICAEDT
51 RVTAQLLSAY GIQGRLVSVR EHNERQMADK VIGFLSDGLV VAQVSDAGTP
101 AVCDPGAKLA RRVREAGFKV VPVVGASAVM AALSVAGVAE SDFYFNGFVP
151 PKSGERRKLF AKWVRAAFPV VMFETPHRIG ATLADMAELF PERRLMLARE
201 ITKTFETFLS GTVGEIQTAL AADGNQSRGE MVLVLYPAQD EKHEGLSESA
251 QNAMKILAAE LPTKQAAELA AKITGEGKKA LYDLALSWKN K*
```

## ORF147ng shows homology to a hypothetical E.coli protein:

```
sp | P45528 | YRAL_ECOLI HYPOTHETICAL 31.3 KD PROTEIN IN AGAI-MTR INTERGENIC REGION
10
          (F286)
          >gi|606086 (U18997) ORF_f286 [Escherichia coli]
          >gi|1789535 (AE000395) hypothetical 31.3 kD protein in agai-mtr intergenic region
          [Escherichia coli] Length = 286
           Score = 218 bits (550), Expect = 3e-56
15
           Identities = 128/284 (45%), Positives = 171/284 (60%), Gaps = 4/284 (1%)
          Query: 4
                     KHLQKASDSVVGGTLYVVATPIGNLADITLRALAVLQKADIICAEDTRVTAQLLSAYGIQ 63
                                 G LY+V TPIGNLADIT RAL VLQ D+I AEDTR T LL +GI
                     KQHQSADNSQ--GQLYIVPTPIGNLADITQRALEVLQAVDLIAAEDTRHTGLLLQHFGIN 59
          Sbjct: 2
20
                     GRLVSVREHNERQMADKVIGFLSDGLVVAQVSDAGTPAVCDPGAKLARRVREAGFKVVPV 123
          Query: 64
                      RL ++ +HNE+Q A+ ++ L +G +A VSDAGTP + DPG L R REAG +VVP+
                     ARLFALHDHNEQQKAETLLAKLQEGQNIALVSDAGTPLINDPGYHLVRTCREAGIRVVPL 119
          Sbjct: 60
25
          Query: 124 VGASAVMAALSVAGVAESDFYFNGFVPPKSGERRKLFAKWVRAAFPVVMFETPHRIGATL 183
                      G A + ALS AG+
                                        F + GF+P KS RR
                                                                   ++ +E+ HR+ +L
          Sbjct: 120 PGPCAAITALSAAGLPSDRFCYEGFLPAKSKGRRDALKAIEAEPRTLIFYESTHRLLDSL 179
          Query: 184 ADMAELFPERR-LMLAREITKTFETFLSGTVGEIQTALAADGNQSRGEMVLVLYPAQDEK 242
30
                      D+ + E R ++LARE+TKT+ET
                                                   VGE+ + D N+ +GEMVL++
          Sbjct: 180 EDIVAVLGESRYVVLARELTKTWETIHGAPVGELLAWVKEDENRRKGEMVLIV-EGHKAQ 238
          Query: 243 HEGLSESAQNAMKILAAELPTKQAAELAAKITGEGKKALYDLAL 286
                           Α
                               + +L AELP K+AA LAA+I G K ALY AL
35
          Sbjct: 239 EEDLPADALRTLALLQAELPLKKAAALAAEIHGVKKNALYKYAL 282
```

Based on the computer analysis and the presence of a putative transmembrane domain in the gonococcal protein, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 77

40 The following partial DNA sequence was identified in N. meningitidis <SEQ ID 647>

	1	ATGAAAACAA	CCGACAAACG	GACAACCGAA	ACACACCGCA	AAGCCCCGAA
	51	AACCGGTCGC	ATCCGCTTCT	C.GCTGCTTA	CTTAGCCATA	TGCCTGTCGT
	101	TCGGCATTCT	TCCCCAAGCC	TGGGCGGGAC	ACACTTATTT	CGGCATCAAC
4.5	151	TACCAATACT	ATCGCGACTT	TGCCGAAAAT	AAAGGCAAGT	TTGCAGTCGG
45	201	GGCGAAAGAT	ATTGAGGTTT	ACAACAAAAA	AGGGGAGTTG	GTCGGCAAAT
	251	CAATGACAAA	AGCCCCGATG	ATTGATTTTT	CTGTGGTGTC	GCGTAACGGC
	301	GTGGCGGcAT	TGGTGGGCGt	<b>ATCAATATAT</b>	TGTGAGCGTG	GCACATAACG
	351	GCGGCTATAA	CAACGTTGAT	TTTGGTGCGG	AAGGAAk.AA	tATCCC GAT
<b>5</b> 0	401	CAACAwCGww	TTACTTATAA	AATTGTGAAA	CGGAATAATT	ATAAAGCAGG
50	451	GACTAAAGGC	CATCCTTATG	GCGGCGATTA	TCATATGCCG	CGTTTGCATA
	501	AATWTGTCAC	AGATGCAGAA	CCTGTTGAAA	TGACCAGTTA	TATGGATGGG
1	551	CGGAAATATA	TCGATCAAAA	TAATTACCCT	GACCGTGTTC	GTATTGGGGC
	601	AGGCAGGCAA	TATTGGCGAT	CTGATGAAGA	TGAGCCCAAT	AACCGCGAAA
	651	GTTCATATCA	TATTGCAAGT			
55	701		GGCTC	ACCAATGTTT	ATCTATGATG	CCCAAAAGCA
	751	AAAGTGGTTA	ATTAATGGGG	TATTGCAAAC	GGGCAACCCC	TATATAGGAA
	801	AAAGCAATGG	CTTCCAGCTG	GTTCGTAAAG	ATTGGTTCTA	TGATGAAATC
	851	TTTGCTGGAG	ATACCCATTC	AGTATTCTAC	GAACCACGTC	AAAATGGGAA
	901	ATACTCTTTT	<b>AACGACGATA</b>	ATAATGGCAC	AGGAAAAATC	AATGCCAAAC
					_	

	951	ATGAACACAA	TTCTCTGCCT	AATAGATTAA	AAACACGAAC	CGTTCAATTG
	1001	TTTAATGTTT	CTTTATCCGA	GACAGCAAGA	GAACCTGTTT :	ATCATGCTGC
	1051	AGGTGGTGTC	AACAGTTATC	GACCCAGACT	GAATAATGGA	GAAAATATTT
	1101	ССТТТАТТСА	CGAAGGAAAA	GGCGAATTGA	TACTTACCAG	CAACATCAAT
`5	1151	CAACCTCCTC	GAGGATTATA	TTTCCAAGGA	GATTTTACGG	TCTCGCCTGA
3	1201	AAATAACGAA	ACTTGGCAAG	GCGCGGGCGT	TCATATCAGT	GAAGACAGTA
	1251	VULTUROGUE	CANACTANAC	CCCCTCCCAA	ACGACCGCCT	GTCCAAAATC
		CCGLIACIIG	CCCMC CANAGE	GGCGIGGCAA		0.00.2
	1301	GGCAMAGGCA	CGCIG	//		•••••
••	0101					СВТАВАС
10	2101		******************************	ACCCACAMCA	GCGGCAATGT	CCATCTTCCC
	2151	TGACTGCTTC	ATTGACIAAG	CACACCCCOM	GCCACACTCA	ACCCCAATCT
	2201	GATCACGCTC	ATTTAAATCT	CACAGGGCII	GCCACACICA	CCCACCCAAA
	2251	TAGTGCAAAT	GGCGATACAC	GTTATACAGT	CAGCCACAAC	MARIOCARA MARIOCARCA
	2301	ACGGCAACCK	TAGCCTCGTG	G. SAATGCCC	AAGCAACATT	TAATCAAGCC
15	2351	ACATTAAACG	GCAACACATC	GGCTTCGGGC	AATGCTTCAT	TTAATCTAAG
	2401	CGACCACGCC	GTACAAAACG	GCAGTCTGAC	GCTTTCCGGC	AACGCTAAGG
	2451	CAAACGTAAG	CCATTCCGCA	CTCAACGGTA	ATGTCTCCCT	AGCCGATAAG
	2501	GCAGTATTCC	ATTTTGAAAG	CAGCCGCTTT	ACCGGACAAA	TCAGCGGCGG
	2551	CAagGATACG	GCATTACACT	TAAAAGACAG	CGAATGGACG	CTGCCGTCAg
20	2601	GarCGGAATT	AGGCAATTTA	AACCTTGACA	ACGCCACCAT	TACaCTCAAT
	2651	TCCGCCTATC	GCCACGATGC	GGCAGGGGCG	CAAACCGGCA	GTGCGACAGA
	2701	TGCGCCGCGC	CGCCGTTCGC	GCCGTTCGCG	CCGTTCCCTA	TTATmCGTTA
	2751	CACCGCCAAC	TTCGGTAGAA	TCCCGTTTCA	ACACGCTGAC	GGTAAACGGC
		AAATTGAACG	GTCAGGGAAC	ATTCCGCTTT	ATGTCGGAAC	TCTTCGGCTA
25	2801 2851	CCGCAGCGAC	AAATTGAAGC	TGGCGGAAAG	TTCCGAAGGC	ACTTACACCT
23	2901	TGGCGGTCAA	CAATACCGGC	AACGAACCTG	CAAGCCTCGA	ACAATTGACG
	2951	GTAGTGGAAG	GAAAAGACAA	CAAACCGCTG	TCCGAAAACC	TTAATTTCAC
	3001	CCTGCAAAAC	GAACACGTCG	ATGCAGGCGC	GTGG	
	3001			//		
30	3551			TTAGAC	CGCGTATTTG	CCGAAGACCG
<b>3</b> 0	3601	CCCCAACGCC	GTTTGGACAA	GCGGCATCCG	GGACACCAAA	CACTACCGTT
	3651	CCCDAGATTT	CCGCGCCTAC	CGCCAACAAA	CCGACCTGCG	CCAAATCGGT
	3701	ATCCACAAAA	ACCTCGGCAG	CGGGCGCGTC	GGCATCCTGT	TTTCGCACAA
	3751	CCCCACCCAA	ABCACCTTCC	ACGACGGCAT	CGGCAACTCG	GCACGGCTTG
25	3801	CCCGGACCGAA	CCTTTTCCCC	CAATACGCA	TCGACAGGTT	CTACATCGGC
35		ATCACTCCC	CCTTTTCGGG	TTACCACCC	CAGCCTTTCA	GACGGCATCG
	3851	WICWGTGGGG	CCCCCCCCCC	11AGCAGCBG	ACGGCATTCA	GGCACGATAC
	3901	GAGSINAAAWI	*C~~CC~ <sup>A</sup> Tt	CCCCAMCCAA	CCGCACATCG	CCCCAACCCC
	3951		CARARACCC	. CGGCAICGAA	CGAAAACGTC	NATATCCCCA
40	4001	CTATTTCGTC	TCCAMMCAACCGC	CC-MACCCCC	CGGGCATTA	CCCACATTAT
40	4051	CCCCCGGCCT	CCCCCCARC	CAMMMACCAM	ACGCCTTATT	TCACCCTCTC
	4101	TCATTCAAAC	CGGCGCAACA	CATTICCATC	ACGCCTTATT	DAMACCIGIC
	4151	CTATACCGAT	GCCGCTTCGG	GCAAAGTCCG	AACACGCGIC	AAIACCGCCG
	4201	TATTGGCTCA	GGATTTCGGC	AAAACCCGCA	GTGCGGAATG	CCCCCANACC
	4251	GCCGAAATCA	AAGGTTTCAC	GCTGTCCCTC	CACGCTGCCG	CCGCCAAAGG
45	4301		GAAGCGCAAC	C ACAGCGCGGG	CATCAAATTA	GGCTACCGCT
	4351	GGTAA				
		3 - 4 - 41 i-			T 649. ODE	1~.
	This correspond	as to the amin	no acia sequ	ence SEQ 1	D 040, UKF	17.
	1	MKTTDKRTTE	THRKAPKTG	R IRFXAAYLAI	CLSFGILPQA	WAGHTYFGIN
	51	YQYYRDFAEN	KGKFAVGAKI	) IEVYNKKGEI	. VGKSMTKAPM	IDFSVVSRNG
50	101	VAALVGVQYI	VSVAHNGGY	NVDFGAEGXN	I IXDQXRXTYK	IVKRNNYKAG
	151	TKGHPYGGDY	HMPRLHKXV	C DAEPVEMTS	MDGRKYIDQN	NYPDRVRIGA
	201	GROYWRSDED	EPNNRESSYI	H IAS	GS	PMFIYDAQKQ
	251		GNPYIGKSNO	G FOLVRKDWFY	DEIFAGDTHS	VFYEPRQNGK
	301	YSFNDDNNGT	GKINAKHEHI	N SLPNRLKTR	VOLFNVSLSE	TAREPVYHAA
55	351	GGVNSYRPRI	NNGENISFI	DEGKGELILTS	NINQGAGGLY	FQGDFTVSPE
33	401	NNETWOGAGY	HISEDSTVT	W KVNGVANDRI	L SKIGKGTL	• • • • • • • • •
	101			//		
	701		DKVTA	S LTKTDISGN	V DLADHAHLNI	TGLATLNGNL
	751 751		CHNATONGN	Y SLVXNAOATI	F NOATLNGNTS	ASGNASFNLS
60	801	DENVIONOST	TSCNAKANU	S HSALMCNUS	. ADKAVEHEES	SRFTGQISGG
00	851	טונגע טונגערט אונע אונע	PROTECTED	T. CMINI.DNAT	T TINCAYEND	AGAQTGSATD
						FRFMSELFGY
	901					KPLSENLNFT
	951				_	
~ =	1001	T-MURHA DAG	1 W	/ /		• • • • • • • • • • • • • • • • • • • •
65						t DDMEREDO
	1151					.LDRVFAEDR
	1201	RNAVWTSGI	R DTKHYRSQD	r KAYRQQTDL	K OTCHOKNICS	GRVGILFSHN
	1251	RTENTEDDG	L GNSARLAHG	A VEGOYGIDR	r YIGISAGAGI	SSGSLSDGIG
	1301	XKXRRRVLH	Y GIQARYRAG	r GGFGIEPHI	G ATRYFVOKAL	YRYENVNIAT
70	1351	PGLAFNRYR	A GIKADYSFK	P AQHISITPY	L SLSYTDAAS	KVRTRVNTAV

1401 LAQDFGKTRS AEWGVNAEIK GFTLSLHAAA AKGPQLEAQH SAGIKLGYRW 1451 \*

## Further sequencing analysis revealed the complete nucleotide sequence <SEQ ID 649>:

•	1	ATGAAAACAA	CCGACAAACG	GACAACCGAA	ACACACCGCA	AAGCCCCGAA
5	51	AACCGGCCGC	ATCCGCTTCT	CGCCTGCTTA	CTTAGCCATA	TGCCTGTCGT
	101			TGGGCGGGAC		
	151			TGCCGAAAAT		
	201	GGCGAAAGAT	ATTGAGGTTT	ACAACAAAAA	AGGGGAGTTG	GTCGGCAAAT
	251	CAATGACAAA	AGCCCCGATG	ATTGATTTTT	CTGTGGTGTC	GCGTAACGGC
10	301	GTGGCGGCAT	TGGTGGGCGA	TCAATATATT	GTGAGCGTGG	CACATAACGG
	351	CGGCTATAAC	AACGTTGATT	TTGGTGCGGA	AGGAAGAAAT	CCCGATCAAC
	401	ATCGTTTTAC	TTATAAAATT	GTGAAACGGA	ATAATTATAA	AGCAGGGACT
	451	AAAGGCCATC	CTTATGGCGG	CGATTATCAT	ATGCCGCGTT	TGCATAAATT
1.5	501	TGTCACAGAT	GCAGAACCTG	TTGAAATGAC	CAGTTATATG	GATGGGCGGA
15	551	AATATATCGA	TCAAAATAAT	TACCCTGACC	GTGTTCGTAT	TGGGGCAGGC
	601	AGGCAATATT	GGCGATCTGA	TGAAGATGAG	CCCAATAACC	GCGAAAGTTC
	651	CACADADAGC	GCAAGTGCGT	ATTCTTGGCT	CGTTGGTGGC	AATACCTTTG
•	701 751	ANACAMACCC	ATCAGGTGGT	GGCACAGTCA	ACTTAGGTAG	TGAAAAAATT
20	801	TECETENCE	AUCUUUNAUCU	TTTACCAACA ATGATGCCCA	GGAGGCTCAT	TTGGCGACAG
20	851	ATGGGGGTATT	CONNECCC	AACCCCTATA	MAAGCAAAAG	TGGTTAATTA
	901	CACCTCCTTC	CTAAACAGGC	GTTCTATGAT	CANATCTTTC	CTCCACATAC
	951	CCATTCAGTA	TTCTACGAAC	CACGTCAAAA	TGGGAAATAC	TCTTTTAACC
	1001	ACGATAATAA	TGGCACAGGA	AAAATCAATG	CCAAACATGA	ACACAATTCT
25	1051	CTGCCTAATA	GATTAAAAAC	ACGAACCGTT	CAATTGTTTA	ATGTTTCTTT
•	1101	ATCCGAGACA	GCAAGAGAAC	CTGTTTATCA	TGCTGCAGGT	GGTGTCAACA
	1151	GTTATCGACC	CAGACTGAAT	AATGGAGAAA	ATATTTCCTT	TATTGACGAA
	1201	GGAAAAGGCG	AATTGATACT	TACCAGCAAC	ATCAATCAAG	GTGCTGGAGG
••	1251	ATTATATTTC	CAAGGAGATT	TTACGGTCTC	GCCTGAAAAT	AACGAAACTT
30	1301	GGCAAGGCGC	GGGCGTTCAT	ATCAGTGAAG	ACAGTACCGT	TACTTGGAAA
•	1351	GTAAACGGCG	TGGCAAACGA	CCGCCTGTCC	AAAATCGGCA	AAGGCACGCT
	1401	GCACGTTCAA	GCCAAAGGGG	AAAACCAAGG	CTCGATCAGC	GTGGGCGACG
•	1451 1501	GTACAGTCAT	TTTGGATCAG	CAGGCAGACG	ATAAAGGCAA	AAAACAAGCC
35	1551	CCATAATCAC	TUGGUTTGGT	CAGCGGCAGG ACAAACTCTA	GGTACGGTGC	AACTGAATGC
JJ	1601	CUMINATURG	AAACCCCCAT	TCGCTTTCGT	TTTCGGCTTT	CGCGGCGGAC
	1651	GATGAAGGGG	CCATCATTCT	CAACCACAAT	CARCOGIAI	AATCCACCCC
	1701	TACCATTACA	GGCAATAAAG	ATATTGCTAC	AACCGGCAAT	AACAACACCT
	1751	TGGATAGCAA	AAAAGAAATT	GCCTACAACG	GTTGGTTTGG	CGAGAAAGAT
40	1801	ACGACCAAAA	CGAACGGGCG	GCTCAACCTT	GTTTACCAGC	CCGCCGCAGA
	1851	AGACCGCACC	CTGCTGCTTT	CCGGCGGAAC	AAATTTAAAC	GGCAACATCA
	1901	CGCAAACAAA	CGGCAAACTG	TTTTTCAGCG	GCAGACCAAC	ACCGCACGCC
	1951	TACAATCATT	TAAACGACCA	TTGGTCGCAA	AAAGAGGGCA	TTCCTCGCGG
45	2001	GGAAATCGTG	TGGGACAACG	ACTGGATCAA	CCGCACATTT	AAAGCGGAAA
45	2051	ACTTCCAAAT	TAAAGGCGGA	CAGGCGGTGG	TTTCCCGCAA	TGTTGCCAAA
	2101	GTGAAAGGCG	ATTGGCATTT	GAGCAATCAC	GCCCAAGCAG	TTTTTGGTGT
•	2151 2201	TCACCACCGCAT	CAAAGCCACA	CAATCTGTAC	ACGTTCGGAC	TGGACGGGTC
	2251	TUACAAATTG	TGTCGAAAAA	ACCATTACCG	ACGATAAAGT	GATTGCTTCA
50	2301	TIGACIAAGA	ACACCCCTTC	CGGCAATGTC CCACACTCAA	GATCTTGCCG	ATCACGCTCA
50	2351	GCGATACACG	TTATACACTC	AGCCACAACG	CCACCCAAAA	CCCCAACCUE
	2401			AGCAACATTT		
	2451	CAACACATCG	GCTTCGGGCA	ATGCTTCATT	TAATCTAACC	CATTAAACGG
	2501	TACAAAACGG	CAGTCTGACG	CTTTCCGGCA	ACGCTAAGGC	AAACGTAAGC
55	2551	CATTCCGCAC	TCAACGGTAA	TGTCTCCCTA	GCCGATAAGG	CAGTATTCCA
	2601	TTTTGAAAGC	AGCCGCTTTA	CCGGACAAAT	CAGCGGCGGC	AAGGATACGG
	2651	CATTACACTT	AAAAGACAGC	GAATGGACGC	TGCCGTCAGG	CACGGAATTA
	2701	GGCAATTTAA	ACCTTGACAA	CGCCACCATT	ACACTCAATT	CCGCCTATCG
60	2751	CCACGATGCG	GCAGGGGCGC	AAACCGGCAG	TGCGACAGAT	GCGCCGCGCC
60	2801	GCCGTTCGCG	CCGTTCGCGC	CGTTCCCTAT	TATCCGTTAC	ACCGCCAACT
	2851	TCGGTAGAAT	CCCGTTTCAA	CACGCTGACG	GTAAACGGCA	AATTGAACGG
	2901	TCAGGGAACA	TTCCGCTTTA	TGTCGGAACT	CTTCGGCTAC	CGCAGCGACA
	2951	AATTGAAGCT	GGCGGAAAGT	TCCGAAGGCA	CTTACACCTT	GGCGGTCAAC
65	3001	AATACCGGCA	ACGAACCTGC	AAGCCTCGAA	CAATTGACGG	TAGTGGAAGG
<del>0</del> 5	3051	AAAAGACAAC	MAACCGCTGT	CCGAAAACCT	TAATTTCACC	CTGCAAAACG
•	3101 3151	CACTOCCCC	TGCCAMA AMOO	TGGCGTTACC	AACTCATCCG	CAAAGACGGC
	3201	GWG1 I CCGCC	TOCKTAMICC	GGTCAAAGAA AACAGGCGGA	AAAAGCTTT	CCCCARACCT
	3251	TTGACCCCC	GATTGCGGCC	GGGCGCGATG	CCCTCCAAAA	CACACAAAGCC
70	3301	GTTGCCGAAC	CGGCCCGGCA	GGCAGGCGAIG	CADADTETCE	GCATTATGCA
		21100001210	3000000	DUDUNG	-WWW. GICG	CONTINICA

	3351	GGCGGAGGAA	GAGAAAAAAC	GGGTGCAGGC	GGATAAAGAC	ACCGCCTTGG
	3401	CGAAACAGCG	CGAAGCGGAA	ACCCGGCCGG	CTACCACCGC	CTTCCCCCGC
	3451	GCCCGCCGCG	CCCGCCGGGA	TTTGCCGCAA	CTGCAACCCC	AACCGCAGCC
	3501	CCAACCGCAG	CGCGACCTGA	TCAGCCGTTA	TGCCAATAGC	GGTTTGAGTG
5	3551	AATTTTCCGC	CACGCTCAAC	AGCGTTTTCG	CCGTACAGGA	CGAATTAGAC
-	3601	CGCGTATTTG	CCGAAGACCG	CCGCAACGCC	GTTTGGACAA	GCGGCATCCG
	3651	GGACACCAAA	CACTACCGTT	CGCAAGATTT	CCGCGCCTAC	CGCCAACAAA
	3701	CCGACCTGCG	CCAAATCGGT	ATGCAGAAAA	ACCTCGGCAG	CGGGCGCGTC
	3751	GGCATCCTGT	TTTCGCACAA	CCGGACCGAA	AACACCTTCG	ACGACGGCAT
10	3801	CGGCAACTCG	GCACGGCTTG	CCCACGGCGC	CGTTTTCGGG	CAATACGGCA
	3851	TCGACAGGTT	CTACATCGGC	ATCAGCGCGG	GCGCGGGTTT	TAGCAGCGGC
	3901	AGCCTTTCAG	ACGGCATCGG	AGGCAAAATC	CGCCGCCGCG	TGCTGCATTA
	3951	CGGCATTCAG	GCACGATACC	GCGCCGGTTT	CGGCGGATTC	GGCATCGAAC
	4001	CGCACATCGG	CGCAACGCGC	TATTTCGTCC	AAAAAGCGGA	TTACCGCTAC
15	4051	GAAAACGTCA	ATATCGCCAC	CCCCGGCCTT	GCATTCAACC	GCTACCGCGC
	4101	GGGCATTAAG	GCAGATTATT	CATTCAAACC	GGCGCAACAC	ATTTCCATCA
	4151	CGCCTTATTT	GAGCCTGTCC	TATACCGATG	CCGCTTCGGG	CAAAGTCCGA
	4201	ACACGCGTCA	ATACCGCCGT	ATTGGCTCAG	GATTTCGGCA	AAACCCGCAG
	4251	TGCGGAATGG	GGCGTAAACG	CCGAAATCAA	AGGTTTCACG	CTGTCCCTCC
20	4301	ACGCTGCCGC	CGCCAAAGGC	CCGCAACTGG	AAGCGCAACA	CAGCGCGGGC
	4351	ATCAAATTAG	GCTACCGCTG	GTAA		

This corresponds to the amino acid sequence <SEQ ID 650; ORF1-1>:

```
1 MKTTDKRTTE THRKAPKTGR IRFSPAYLAI CLSFGILPQA WAGHTYFGIN
                     YQYYRDFAEN KGKFAVGAKD IEVYNKKGEL VGKSMTKAPM IDFSVVSRNG
                     VAALVGDQYI VSVAHNGGYN NVDFGAEGRN PDQHRFTYKI VKRNNYKAGT
25
                101
                     KGHPYGGDYH MPRLHKFVTD AEPVEMTSYM DGRKYIDQNN YPDRVRIGAG
                151
                     ROYWRSDEDE PNNRESSYHI ASAYSWLVGG NTFAQNGSGG GTVNLGSEKI
                201
                     KHSPYGFLPT GGSFGDSGSP MFIYDAQKQK WLINGVLQTG NPYIGKSNGF
                251
                     QLVRKDWFYD EIFAGDTHSV FYEPRONGKY SFNDDNNGTG KINAKHEHNS
                301
                     LPNRLKTRTV QLFNVSLSET AREPVYHAAG GVNSYRPRLN NGENISFIDE
30
                351
                401
                     GKGELILTSN INOGAGGLYF QGDFTVSPEN NETWQGAGVH ISEDSTVTWK
                     VNGVANDRLS KIGKGTLHVQ AKGENQGSIS VGDGTVILDQ QADDKGKKQA
                451
                     FSEIGLVSGR GTVQLNADNQ FNPDKLYFGF RGGRLDLNGH SLSFHRIQNT
                501
                     DEGAMIVNHN QDKESTVTIT GNKDIATTGN NNSLDSKKEI AYNGWFGEKD
                551
35
                     TTKTNGRLNL VYQPAAEDRT LLLSGGTNLN GNITQTNGKL FFSGRPTPHA
                601
                     YNHLNDHWSQ KEGIPRGEIV WDNDWINRTF KAENFQIKGG QAVVSRNVAK
                651
                     VKGDWHLSNH AQAVFGVAPH QSHTICTRSD WTGLTNCVEK TITDDKVIAS
                701
                     LTKTDISGNV DLADHAHLNL TGLATLNGNL SANGDTRYTV SHNATQNGNL
                751
                     SLVGNAQATF NQATLNGNTS ASGNASFNLS DHAVQNGSLT LSGNAKANVS
                801
                     HSALNGNVSL ADKAVFHFES SRFTGQISGG KDTALHLKDS EWTLPSGTEL
40
                851
                     GNLNLDNATI TLNSAYRHDA AGAQTGSATD APRRRSRRSR RSLLSVTPPT
                901
                     SVESRFNTLT VNGKLNGQGT FRFMSELFGY RSDKLKLAES SEGTYTLAVN
                951
                     NTGNEPASLE QLTVVEGKDN KPLSENLNFT LQNEHVDAGA WRYQLIRKDG
               1001
                     EFRLHNPVKE QELSDKLGKA EAKKQAEKDN AQSLDALIAA GRDAVEKTES
               1051
                     VAEPARQAGG ENVGIMQAEE EKKRVQADKD TALAKQREAE TRPATTAFPR
45
               1101
                     ARRARRDLPQ LQPQPQPQPQ RDLISRYANS GLSEFSATLN SVFAVQDELD
               1151
                     RVFAEDRRNA VWTSGIRDTK HYRSQDFRAY RQQTDLRQIG MQKNLGSGRV
               1201
               1251
                     GILFSHNRTE NTFDDGIGNS ARLAHGAVFG QYGIDRFYIG ISAGAGFSSG
                     SLSDGIGGKI RRRVLHYGIQ ARYRAGFGGF GIEPHIGATR YFVQKADYRY
               1301
50
               1351
                     ENVNIATPGL AFNRYRAGIK ADYSFKPAQH ISITPYLSLS YTDAASGKVR
                     TRVNTAVLAQ DFGKTRSAEW GVNAEIKGFT LSLHAAAAKG PQLEAQHSAG
               1401
               1451
                     IKLGYRW*
```

Computer analysis of these sequences gave the following results:

## Homology with a predicted ORF from N. meningitidis (strain A)

ORF1 shows 57.8% identity over a 1456aa overlap with an ORF (ORF1a) from strain A of N. meningitidis:

		10	20	30	40	50	60
	orfl.pep	MKTTDKRTTETHRK	APKTGRIRF7	KAAYLAICLSF	GILPQAWAGE	ITYFGINYQYY	RDFAEN
	• •	1111111111111111	$\Pi\Pi\Pi\Pi\Pi\Pi$	1111111111		11111111111	
60	orf1a	MKTTDKRTTETHRK	APKTGRIRES	SPAYLAICLSF	GILPQAWAGH	TYFGINYQYY	RDFAEN
		10	20	30	40	50	60
		70	80	90	100	110	120
	orf1.pep	KGKFAVGAKDIEVY	NKKGELVGK:	SMTKAPMIDFS	VVSRNGVAAI	VGVQYIVSVI	HNGGYN

·	orfla	!          KGKFAVGAKE 70	IEVYNKKG					
5	orfl.pep	130 NVDFGAEGXN	) 1 IIXDQXRXT	40 YKIVKRNN	150 IYKAGTKGH	160 PYGGDYHMPF	170 RLHKXVTDA	180 AEPVEMTSY
	orfla	NVDFGAEGXN	-PDQHRFS	YQIVKRNN 140	IYKPDNS-H 150	PYNGDXHMPF	RLHKFVTDA 170	EPVEMTSD
10		190		00	210	160	170	,
	orfl.pep	MDGRKYIDQN	NYPDRVRI	GAGRQYWE	SDEDEP			NN
15	orfla	MRGNTYSDKE		GSGHHYWF		LSYSGAWLIC 220	GGNTHMQGW 2,30	GNNGVXSL
	orf1.pep	220 RESSYF	IIA	SGSPN	230 MFIYDAQKQ	240 KWLINGVLQT	250 IGNPYIGKS	260 SNGFQLVRK
20	orfla	SGDVRHANDY	GPMPÍAGA 250	AGDSGSPN 260	IFIYDKTNN 270	KWLLNGVLO	rgypysgre 290	ENGFQLIRK
		270	28			300	310	320
25	orfl.pep	DWFYDEIFAG	1111:1:1	11:11::1	:::	1:: :1:	1 1:11:	::!!:!!:
	orfla	DWFYDDIYRO	SDTHTVXFE 310	PRSNGHFS 320	FTSNNNGT 330	GTVTETNEK 340		/QTVRLFDE 50
30	orfl.pep	330 SLSETAREP	34		350	360	370	380
30	orfla	:   :    SLNETDKEP	1 11111	:111111	H11:1111	1:1:111:	: [] [] [] []	1111:1111
		360	370	380	390			110
35	orfl.pep	390 VSPENNETW(	40 GAGVHISE		410 MGVANDRI	420 SKIGKGTL-	430	
	orfla	VSPENNETWO		1111111		1111111		SSISVGDGT
40		420	430	440	450			470
	orfl.pep							
45	orfla	VILDQQADD	KGKKQAFSE				GFRGGRLD!	LNGHSLSFH
45		480	490	500	510	520	) !	530
	orfl.pep							
50	orfla	RIQNTDEGAN 540	MIXXHNATT 550	TSTVTITO 560	GNESITOPS 570			WFGEKDTTK 590
	orfl.pep							
55	orfla	TNCPINIVY	7022E00TY	T T SCCTINI	NCNTTOTA	CVI PPCCDD	BDURVAUL.	
	Ollia	600	610	620	630	IGKLFFSGRP 64		650
60	orf1.pep							
	orfla	IPOGEIVWD	NDWIXRTFK	(AENFHIO	GGOAVISRN	IVAKVEGDXH	LSNHAOAV	FGVAPHOSH
		660	670	680	690			710
65	orfl.pep		440 XX		450 ASLTKTDIS	460 GNVDLADHA	470 HLNLTGLA	480 TLNGNLSAN
	orfla		: 11	: 111	1111111		1:11	:1 1111111
70		720	730	740	750			770
	orfl.pep	490 GDTRYTVSH	500 NATQNGNXS		510 FFNQATLNO	520 INTSASGNAS	530 FNLSDHAV	540 QNGSLTLSG

-363-

	orfla	
5	orfl.pep	550 560 570 580 590 600 NAKANVSHSALNGNVSLADKAVFHFESSRFTGQISGGKDTALHLKDSEWTLPSGXELGNL
10	orfla	NAKANVSHSALNGNVSLADKAVFHFENSRFTGQLSGSKXTALHLKDSEWTLPSGTELGNL 840 850 860 870 880 890
	orfl.pep	610 620 630 640 650 660  NLDNATITLNSAYRHDAAGAQTGSATDAPRRSRSRSLLXVTPPTSVESRFNTLTVNG
15		900 910 920 930 940 950 670 680 690 700 710 720
20	orfl.pep	KLNGQGTFRFMSELFGYRSDKLKLAESSEGTYTLAVNNTGNEPASLEQLTVVEGKDNKPL
25	orfl.pep	730 740 750  SENLNFTLQNEHVDAGAW
30	orfl.pep	1020 1030 1040 1050 1060 1070
	orfla	LDALIAAGRDAAEKTESVAEPARXAGGENVGIMQAEEEKKRVQADKDSALAKQREAETRP 1080 1090 1100 1110 1120 1130
35	orfl.pep	760 LDR
40	orfla	XTTAFPRARXARRDLPQPQPQPQPQPQPQRDLXSRYANSGLSEFSATLNSVFAVQDELDR 1140 1150 1160 1170 1180 1190
	orfl.pep orfla	770 780 790 800 810 820  VFAEDRRNAVWTSGIRDTKHYRSQDFRAYRQQTDLRQIGMQKNLGSGRVGILFSHNRTEN
45		1200 1210 1220 1230 1240 1250 830 840 850 860 870 880
50	orfl.pep	TFDDGIGNSARLAHGAVFGQYGIDRFYIGISAGAGFSSGSLSDGIGXKXRRVLHYGIQA :
55	orfl.pep	890 900 910 920 930 940 RYRAGFGGFGIEPHIGATRYFVQKADYRYENVNIATPGLAFNRYRAGIKADYSFKPAQHI
60	orfl.pep	950 960 970 980 990 1000 SITPYLSLSYTDAASGKVRTRVNTAVLAQDFGKTRSAEWGVNAEIKGFTLSLHAAAAKGP
65	orfl.pep orfla	1010 1020 QLEAQHSAGIKLGYRWX

70 The complete length ORF1a nucleotide sequence <SEQ ID 651> is:

	1	,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,	0001011100	C1C11000		
	1 51	ATGAMAGCAA	AMCCCCMMCM	GACAACCGAA	ACACACCGCA	AAGCCCCGAA
	101	TCCCC NTTCT	MCCCGCTTCT	CGCCTGCTTA	CTTAGCCATA	TGCCTGTCGT
	151	TACCAATACT	AMCCCCAAGCT	TGGGCGGGAC	ACACTTATTT	CGGCATCAAC
5	201	CCCCAAIACI	ATCGCGACTT	TGCCGAAAAT	AAAGGCAAGT	TTGCAGTCGG
. , ,	251	CAAMCACAAA	ATTGAGGTNT	ACAACAAAA	AGGGGAGTTG	GTCGGCAAAT
	301	CTCCCCCCAT	MCCCCCGATG	ATTGATTTTT	CTGTGGTGTC	GCGTAACGGC
	351	GIGGCGGCAI	A A COMMON MAN	TCAATATATT	GTGAGCGTGG	CACATAACGG
	401	ACCOMMMM	MACGITGATT	TTGGTGCGGA	AGGAAGNAAT	CCCGATCAGC
10		MCCGITTITC	TTACCAAATT	GTGAAAAGAA	ATAATTATAA	GCCTGACAAT
10	451	TCACACCCTT	ACAACGGCGA	TTANCATATG	CCGCGTTTGC	ATAAATTTGT
	501	CACAGATGCA	GAACCTGTCG	AAATGACGAG	TGACATGAGG	GGGAATACCT
	551	ATTCCGATAA	AGAAAAATAT	CCCGAGCGTG	TCCGCATCGG	CTCAGGACAC
	601	CACTATTGGC	GTTATGATGA	TGACAAACAC	GGCGATTTAT	CCTACTCCGG
16	651	CGCATGGTTA	ATTGGCGGCA	ATACACATAT	GCAGGGTTGG	GGAAATAATG
15	701	GCGTANTTAG	TTTGAGCGGC	GATGTGCGCC	ATGCCAACGA	CTATGGCCCT
	751	ATGCCGATTG	CAGGTGCGGC	AGGCGACAGC	GGTTCGCCAA	TGTTTATTTA
	801	TGACAAAACA	AACAATAAAT	GGCTGCTCAA	CGGAGTTTTA	CAAACCGGCT
	851	ACCCTTATTC	CGGCAGGGAA	AACGGTTTCC	AGCTGATACG	CAAAGATTGG
00	901	TTCTACGATG	ACATTTACAG	AGGCGATACA	CATACCGTCT	NTTTTGAACC
20	951	GCGCAGTAAC	GGACATTTTT	CCTTTACATC	CAACAACAAC	GGTACGGGTA
	1001	CGGTAACAGA	AACCAACGAA	AAGGTNTCCA	ATCCAAAGCT	TAAAGTACAG
	1051	ACAGTCCGAC	TGTTTGACGA	ATCTTTGAAT	GAAACTGATA	AAGAACCAGT
	1101	TTACGCGGCA	GGGGGTGTTA	ATCAGTACCG	TCCAAGGTTA	AACAACGGTG
	1151	AAAACCTTTC	TTTTATCGAT	TACGGCAACG	GCAAACTCAT	CTTATCAAAC
25	1201	AACATCAACC	AAGGCGCGGG	CGGTTTGTAT	TTTGAAGGTG	ATTTTACCCT
	1251	CTCGCCTGAA	AACAACGAAA	CGTGGCAAGG	CGCGGGCGTT	CATATCACTC
	1301	AAGACAGTAC	CGTTACTTGG	AAAGTAAACG	GCGTGGCAAA	CGACCGCCTG
•	1351	TCCAAAATCG	GCAAAGGCAC	GCTGCACGTT	CAAGCCAAAG	GGGAAAACCA
	1401	AGGCTCGATC	AGCGTGGGCG	ACGGTACAGT	CATTTTGGAT	CAGCAGGCAG
30	1451	ACGATAAAGG	CAAAAAACAA	GCCTTTAGTG	AAATCGGCTT	CNTCACCCCC
•	1501	AGGGGTACGG	TGCAACTGAA	TGCCGATAAT	CAGTTCAACC	CCCACAAACT
	1551	CTATTTCGGC	TTTCGCGGCG	GACGTTTGGA	TTTAAACGGG	CATTCCCTTT
•	1601	CGTTCCACCG	TATTCAAAAT	ACCGATGAAG	GGGCGATGAT	TCNCNATCAT
	1651	AATGCCACAA	CAACATCCAC	CGTTACCATT	ACAGGGAATG	AAACTATTAC
35	1701	ACAACCGAGT	GGTAAGAATA	TCAATAGACT	TAATTACAGC	AAAGAAATTG
	1751	CCTACAACGG	TTGGTTTGGC	GAGAAAGATA	CCACCAAAAC	CANCECCCC
•	1801	CTCAACCTTG	TTTACCAGCC	CGCCGCAGAA	GACCGCACCC	NGCTCCTTTC
	1851	CGGCGGAACA	AATTTAAACG	GCAACATCAC	GCDDDCCACC	CCCNANCTCT
	1901	TTTTCAGCGG	CAGACCGACA	CCGCACGCCT	DCD DTCDTTT	ACCARACIGI
40	1951	TGGTCAAAAA	TGGAAGGTAT	CCCACAAGGA	CANATCATT	CCCACAACCA
	2001	CTGGATCNAC	CGCACGTTTA	AAGCGGAAAA	TTTTCCTGTGT	CACCCCCCC
	2051	AGGCGGTGAT	TTCCCGCAAT	GTTGCCAAAG	TEGAACECEA	TTCNCATTE
	2101	AGCAATCACG	CCCAAGCAGT	TTTTGGTGTC	COLCCCATC	DARCCORMAG
	2151	AATCTGTACA	CGTTCGGACT	GGACNGGTCT	CACAAATTCT	CTCCAANAA
45	2201	NCATTACCGA	CGATAAAGTG	ATTGCTTCAT	TCACTAACAC	NGACNTNAGC
	2251	GGCANTGTNA	GNCTNNCCNA	TNACGNTNNT	TNIANANCTON	CNCCCOMBCC
	2301	NNCACTNAAN	GGCAATCTTA	GTGCAAATGG	CCVAVCVCCM	CNGGGCNTGC
	2351	GCCACAACGC	CACCCAAAAC	GGCAACCTTA	CCCTCCTCCT	CARROCCCA
	2401	GCAACATTTA	ATCAACCCAC	ATTAAACGGC	OCCICGIGGG	NAMEGOCCAA
50	2451	TGCTTCATTT	AATCTAACCA	ACAACGCCGC	ACANANCCCC	NITCGGGCAA
	2501	TTTCCCACAA	CCCTAACCCA	AACGTAAGCC	ACAAAACGGC	AGTCTGACGC
	2551	GTCTCCCTAG	CCCATAACCC	AGTATTCCAT	MUMCARAGE	CAACGGCAAT
	2601	CCCACAACTC	ACCCCCACCA	AGGANACAGC	TITGAAAACA	GCCGCTTTAC
	2651	AATGGACGCT	CCCCTCACCC	ACGGAATTAG	CCAADDDDAA	AAAGACAGCG
55	2701	GCCACCATTA	CACTCAATTC	CGCCTATCGC	CACCAMO	CCTTGACAAC
	2751	AACCGCCACN	CTCTCACACA	CGCCGCGCCG	CACGATGCTG	CAGGCGCGCA
	2801	TATCCCCCTTAC	DCCCCCD DCT	TCGGTAGAAT	CCGTTCGCGC	CGTTCCCTAT
	2851	GTAAACGGCA	ACCOCCARCI	TCAAGGAACA	MMCCCCMTTCAA	CACGCTGACG
	2901	CTTCCCCTAC	CCAACCCACA	AATTGAAGCT	TTCCGCTTTA	TGTCGGAACT
60	2951	CTTCGGCTAC	CCCCCTCAAC	AATIGAAGCT	GGCGGAAAGT	TCCGAAGGNA
00	3001	CITACACCIT	TACTOCO I CAME	AATACCGGCA	ACGAACCCGT	AAGCCTCGAT
	3051	TAATTGACGG	CHCCANANCC	GAAAGACAAC	AAACCGCTGT	CCGAAAACCT
	3101	AACTICACC	CIGCAAAACG	AACACGTCGA	TGCCGGCGCG	TGGCGTTACC
		AACTCATCCG	CAAAGACGGC	GAGTTCCGCC	TGCATAATCC	GGTCAAAGAA
65	3151	CAAGAGCTTT	CCGACAAACT	CGGCAAGGCA	GAAGCCAAAA	AACAGGCGGA
0.5	3201	MANAGACAAC	GCGCAAAGCC	TTGACGCGCT	GATTGCGGCC	GGGCGCGATG
	3251	CCGCCGAAAA	GACAGAAAGC	GTTGCCGAAC	CGGCCCGGCN	GGCAGGCGGG
	3301	GAAAATGTCG	GCATTATGCA	GGCGGAGGAA	GAGAAAAAAC	GGGTGCAGGC
	3351	GGATAAAGAC	AGCGCNTTGG	CGAAACAGCG	CGAAGCGGAA	ACCCGGCCGG
70	3401	NTACCACCGC	CTTCCCCCGC	GCCCGCNGCG	CCCGCCGGGA	TTTGCCGCAA
70	3451	CCGCAGCCCC	AACCGCAACC	TCAACCCCAA	CCGCAGCGCG	ACCTGATNAG
	3501	CCGTTATGCC	AATAGCGGTT	TGAGTGAATT	TTCCGCCACG	CTCAACAGCG
	3551	TTTTCGCCGT	ACAGGACGAA	TTGGACCGCG	TGTTTGCCGA	AGACCGCCGC
•						

3651 AGATTTCCGC GCCTACCGCC AACAAACCGA CCTGCGCCA	
	0.000.000.000
3701 AGAAAAACCT CGGCAGCGGG CGCGTCGGCA TCCTGTTTT	C GCACAACCGG
3751 ACCGAAAACA NCTTCGACGA CGGCATCGGC AACTCGGCA	C GGCTTGCCCA
5 3801 CGGCGCCGTT TTCGGGCAAT ACGGCATCGG CAGGTTCGA	C ATCGGCATCA
3851 GCACGGCGC GGGTTTTAGC AGCGGCANTC TNTCAGACG	G CATCGGAGGC
3901 AAAATCCGCC GCCGCGTGCT GCATTACGGC ATTCAGGCA	C GATACCGCGC
3951 CGGTTTCGGC GGATTCGGCA TCGAACCGTA CATCGGCGC	A ACGCGCTATT
4001 TCGTCCAAAA AGCGGATTAC CGCTACGAAA ACGTCAATA	T CGCCACCCC
10 4051 GGTCTTGCGT TCAACCGNTA CCGNGCGGGC ATTAAGGCA	G ATTATTCATT
4101 CAAACCGGCG CAACACATNT CCATCACNCC TTATTTNAG	C CTGTCCTATA
4151 CCGATGCCGC TTCGGGCAAA GTCCGAACAC GCGTCAATA	C CGCNGTATTG
4201 GCTCAGGATT TCGGCAAAAC CCGCAGTGCG GAATGGGGC	G TAAACGCCGA
4251 AATCAAAGGT TTCACGCTGT CCNTCCACGC TGCCGCCGC	C AAAGGNCCGC
15 4301 AACTGGAAGC GCAACACAGC GCGGGCATCA AATTAGGCT	A CCGCTGGTAA

## This encodes a protein having amino acid sequence <SEQ ID 652>:

	. 1				CLSFGILPQA	
	51	YQYYRDFAEN	KGKFAVGAKD	IEVYNKKGEL	VGKSMTKAPM	IDFSVVSRNG
	101				PDQHRFSYQI	
20	151				GNTYSDKEKY	
	201				GNNGVXSLSG	
	251				QTGYPYSGRE	
	301				GTGTVTETNE	
	351		ETDKEPVYAA	GGVNQYRPRL	NNGENLSFID	YGNGKLILSN
25	401				HISEDSTVTW	
	451				QQADDKGKKQ	
	501	RGTVQLNADN	QFNPDKLYFG	FRGGRLDLNG	HSLSFHRIQN	TDEGAMIXXH
	551				KEIAYNGWFG	
	601	LNLVYQPAAE	DRTXLLSGGT	NLNGNITQTN	GKLFFSGRPT	PHAYNHLGSG
30	651	WSKMEGIPQG	EIVWDNDWIX	RTFKAENFHI	QGGQAVISRN	VAKVEGDXHL
	701	SNHAQAVFGV	APHQSHTICT	RSDWTGLTNC	VEXXITDDKV	IASLTKTDXS
	751	GXVXLXXXXX	XXLXGXAXLX	GNLSANGDTR	YTVSHNATQN	GNLSLVGNAQ
	801	-			SLTLSDNAKA	
	851				KDSEWTLPSG	
35	901				RSLLSVTPPT	
	951				SEGTYTLAVN	
	1001				WRYQLIRKDG	
	1051	QELSDKLGKA	EAKKQAEKDN	AQSLDALIAA	GRDAAEKTES	VAEPARXAGG
40	1101				TRPXTTAFPR	
40	1151	- ~ - ~ - ~ - ~			LNSVFAVQDE	
	1201				IGMQKNLGSG	
	1251				IGISTGAGFS	
	1301				TRYFVQKADY	
4.5	1351				LSYTDAASGK	
45	1401	aqdfgktrsa	EWGVNAEIKG	FTLSXHAAAA	KGPQLEAQHS	AGIKLGYRW*

A transmembrane region is underlined.

## ORF1-1 shows 86.3% identity over a 1462aa overlap with ORF1a:

50	orfla.pep	10 MKTTDKRTTETHRE		HHHHHH	1111111111	111111111	111111
		70	80	90	100	110	120
55	orfla.pep	KGKFAVGAKDIEVY	NKKGELVGKS				
		1111111111111111		1111111111	111111111	1111111111	111111
	orf1-1	KGKFAVGAKDIEV	NKKGELVGKS	MTKAPMIDFS	VVSRNGVAAI	VGDQYIVSVA	HNGGYN
		70	80	90	100	110	120
60		130	140	150	160	170	179
	orfla.pep	NVDFGAEGXNPDQF					
		111111111111111			1:11 1111	1111111111	IIIII
	orf1-1	NVDFGAEGRNPDQF	IRFTYKIVKRN			HKFVTDAEPV	EMTSYM
		130	140	150	160	170	180
65							

	orfla.pep	180 190 200 210 220 230 RGNTYSDKEKYPERVRIGSGHHYWRYDDDKHGDLSYSGAWLIGGNTHMQGWGNN
5	orfl-1	:::  :     : ::      : :::                : :      ::::  DGRKYIDQNNYPDRVRIGAGRQYWRSDEDEPNNRESSYHIASAYSWLVGGNTFAQNGSGG 190 200 210 220 230 240
	orfla.pep	240 250 260 270 280 290 GVXSLSGD-VRHANDYGPMPIAGAAGDSGSPMFIYDKTNNKWLLNGVLQTGYPYSGRENG ::::::::::::::::::::::::::::::::::::
10	orfl-1	GTVNLGSEKIKHS-PYGFLPTGGSFGDSGSPMFIYDAOKOKWLINGVLQTGNPYIGKSNG 250 260 270 280 290
15	orfla.pep	300 310 320 330 340 350 FQLIRKDWFYDDIYRGDTHTVXFEPRSNGHFSFTSNNNGTGTVTETNEKVSNP-KLKVQT
	orf1-1	FQLVRKDWFYDEIFAGDTHSVFYEPRONGKYSFNDDNNGTGKINAKHEHNSLPNRLKTRT 300 310 320 330 340 350
20	orfla.pep	360 370 380 390 400 410 VRLFDESLNETDKEPVY-AAGGVNQYRPRLNNGENLSFIDYGNGKLILSNNINQGAGGLY  :  :   :  :   :
25	orfl-1	VQLFNVSLSETAREPVYHAAGGVNSYRPRLNNGENISFIDEGKGELILTSNINQGAGGLY 360 370 380 390 400 410
	orfla.pep	420 430 440 450 460 470 FEGDFTVSPENNETWQGAGVHISEDSTVTWKVNGVANDRLSKIGKGTLHVQAKGENQGSI
30	orf1-1	420 430 440 450 460 470
:	orfla.pep	480 490 500 510 520 530 SVGDGTVILDQQADDKGKKQAFSEIGLXSGRGTVQLNADNQFNPDKLYFGFRGGRLDLNG
35		SVGDGTVILDQQADDKGKKQAFSEIGLVSGRGTVQLNADNQFNPDKLYFGFRGGRLDLNG 480 490 500 510 520 530 540 550 560 570 580 500
40	orfla.pep	HSLSFHRIQNTDEGAMIXXHNATTTSTVTITGNESITQPSGKNINRLNYSKEIAYNGWFG
		HSLSFHRIQNTDEGAMIVNHNQDKESTVTITGNKDIAT-TGNN-NSLDSKKEIAYNGWFG 540 550 560 570 580 590 600 610 620 630 640 650
45	orfla.pep	EKDTTKTNGRLNLVYQPAAEDRTXLLSGGTNLNGNITQTNGKLFFSGRPTPHAYNHLGSG
	0111 1	EKDTTKTNGRLNLVYQPAAEDRTLLLSGGTNLNGNITQTNGKLFFSGRPTPHAYNHLNDH 600 610 620 630 640 650 660 670 680 690 700 710
50	orfla.pep	WSKMEGIPQGEIVWDNDWIXRTFKAENFHIQGGQAVISRNVAKVEGDXHLSNHAQAVFGV
55		WSQKEGIPRGEIVWDNDWINRTFKAENFOIKGGQAVVSRNVAKVKGDWHLSNHAQAVFGV 660 670 680 690 700 710 720 730 740 750 760 770
	orfla.pep	APHOSHTICTRSDWTGLTNCVEXXITDDKVIASLTKTDXSGXVXLXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
60		APHOSHTICTRSDWTGLTNCVEKTITDDKVIASLTKTDISGNVDLADHAHLNLTGLATLN 720 730 740 750 760 770  780 790 800 810 820 820
	orfla.pep	GNLSANGDTRYTVSHNATQNGNLSLVGNAQATFNQATLNGNXSXSGNASFNLSNNAAQNG
65	0111-1	GNLSANGDTRYTVSHNATQNGNLSLVGNAQATFNQATLNGNTSASGNASFNLSDHAVQNG 780 790 800 810 820 830
70	orfla.pep	840 850 860 870 880 890 SLTLSDNAKANVSHSALNGNVSLADKAVFHFENSRFTGQLSGSKXTALHLKDSEWTLPSG
70	orf1-1	SLTLSGNAKANVSHSALNGNVSLADKAVFHFESSRFTGQISGGKDTALHLKDSEWTLPSG 840 850 860 870 880 890

-367-

5	orfla.pep	900 910 920 930 940  TELGNLNLDNATITLNSAYRHDAAGAQTGXVSDTPRRSRRSLLSVTPPTSVESRFN
10	orfla.pep	950 960 970 980 990 1000 TLTVNGKLNXQGTFRFMSELFGYRSDKLKLAESSEGTYTLAVNNTGNEPVSLDQLTVVEG
15	orfla.pep	1010 1020 1030 1040 1050 1060  KDNKPLSENLNFTLQNEHVDAGAWRYQLIRKDGEFRLHNPVKEQELSDKLGKAEAKKQAE
20	orfla.pep	1070 1080 1090 1100 1110 1120  KDNAQSLDALIAAGRDAAEKTESVAEPARXAGGENVGIMQAEEEKKRVQADKDSALAKQR
25	orfla.pep	1130 1140 1150 1160 1170 1180  EAETRPXTTAFPRARXARRDLPQPQPQPQPQPQPQRDLXSRYANSGLSEFSATLNSVFAV
30	orf1-1	1140 1150 1160 1170 1180 1190  1190 1200 1210 1220 1230 1240  QDELDRYFAEDRRNAVWTSXIRXTKHYRSQDFRAYRQQTDLRQIGMQKNLGSGRVGILFS
35	orfl-1	
40	orfla.pep	1250 1260 1270 1280 1290 1300  HNRTENXFDDGIGNSARLAHGAVFGQYGIGRFDIGISTGAGFSSGXLSDGIGGKIRRRVL      :
45	orfla.pep	1310 1320 1330 1340 1350 1360 HYGIQARYRAGFGGFGIEPYIGATRYFVQKADYRYENVNIATPGLAFNRYRAGIKADYSF
50	orfla.pep	1320 1330 1340 1350 1360 1370  1370 1380 1390 1400 1410 1420  KPAQHXSITPYXSLSYTDAASGKVRTRVNTAVLAQDFGKTRSAEWGVNAEIKGFTLSXHA
55	orfl-1	KPAQHISITPYLSLSYTDAASGKVRTRVNTAVLAQDFGKTRSAEWGVNAEIKGFTLSLHA 1380 1390 1400 1410 1420 1430 1430 1440 1450
60	orfla.pep	AAAKGPQLEAQHSAGIKLGYRWX 
UU	Homology with adh	esion and penetration protein hap precursor of H.influenzae (accession number P45387)
	Amino acids 23-4	23 of ORF1 show 59% aa identity with hap protein in 450aa overlap:
	orfl 23	FXAAYLAICLSFGILPQAWAGHTYFGINYQYYRDFAENKGKFAVGAKDIEVYNKKGELVG 82 F +L C+S GI QAWAGHTYFGI+YQYYRDFAENKGKF VGAK+IEVYNK+G+LVG
65	hap 6	FRLNFLTACVSLGIASQAWAGHTYFGIDYQYYRDFAENKGKFTVGAKNIEVYNKEGQLVG 65
	orf1 83	KSMTKAPMIDFSVVSRNGVAALVGVQYIVSVAHNGGYNNVDFGAEGXNIXDQXRXTYKIV 142 SMTKAPMIDFSVVSRNGVAALVG QYIVSVAHNGGYN+VDFGAEG N DQ R TY+IV
	hap 66	

			•	
	orf1		KRNNYKAGTKGHPYGGDYHMPRLHKXVTDAEPVEMTSYMDGRKYIDQNNYPDRVRIGAGR KRNNY+A + HPY GDYHMPRLHK VT+AEPV MT+ MDG+ Y D+ NYP+RVRIG+GR	
5	hap	125	KRNNYQAWERKHPYDGDYHMPRLHKFVTEAEPVGMTTNMDGKVYADRENYPERVRIGSGR	184
	orf1	203	QYWRSDEDEPNNRESSYHIAQYWR+D+DE N SSY+++	222
	hap	185	QYWRTDKDEETNVHSSYYVSGAYRYLTAGNTHTQSGNGNGTVNLSGNVVSPNHYGPLPTG	244
10	orfl	223	SGSPMFIYDAQKQKWLINGVLQTGNPYIGKSNGFQLVRKDWFYDEIFAGDTHSVF SGSPMFIYDA+K++WLIN VLQTG+P+ G+ NGFQL+R++WFY+E+ A DT SVF	277
	hap	245	GSKGDSGSPMFIYDAKKKQWLINAVLQTGHPFFGRGNGFQLIREEWFYNEVLAVDTPSVF	304
15	orfl	278	YEPRQNGKYSFNDDNNGTGKIN-AKHEHNSLPNRLKTRTVQLFNVSLSETAREPVYHA Y P NG YSF +N+GTGK+ + + + + TV+LFN SL++TA+E V A	334
	hap	305	QRYIPPINGHYSFVSNNDGTGKLTLTRPSKDGSKAKSEVGTVKLFNPSLNQTAKEHV-KA	363
	orfl		AGGVNSYRPRLNNGENISFIDEGKGELILTSNINQGAGGLYFQGDFTV-SPENNETWQGAAA G N Y+PR+ G+NI D+GKG L + +NINQGAGGLYF+G+F V +NN TWQGA	
20	hap	364	AAGYNIYQPRMEYGKNIYLGDQGKGTLTIENNINQGAGGLYFEGNFVVKGKQNNITWQGA	423
	orfl	•	GVHISEDSTVTWKVNGVANDRLSKIGKGTL 423 GV I +D+TV WKV+ NDRLSKIG GTL	
	hap		GVSIGQDATVEWKVHNPENDRLSKIGIGTL 453	
25	Amino acids	715-	1011 of ORF1 show 50% aa identity with hap protein in 258aa overla	ap:
	Orfl	41	DTRYTVSHNATQ-NGNXSLVXNAQATFNQ-ATLNGNTSASGNASFNLSDHAVQNGSLTLS DT+ S TQ NG+ +L NA + A LNGN + ++ F LS++A Q G++ LS	98
	hap	733	DTKVINSIPITQINGSINLTNNATVNIHGLAKLNGNVTLIDHSQFTLSNNATQTGNIKLS	792
30	orf1	99	GNAKANVSHSALNGNVSLADKAVFHFESSRFTGQISGGKDTALHLKDSEWTLPSGXELGN : +A A V+++ LNGNV L D A F ++S F QI G KDT + L+++ WT+PS L N	158
	hap	793	THE A V+++ LNGNV L D A F ++S F QI G KDT + L+++ WT+PS L N NHANATVNNATLNGNVHLTDSAQFSLKNSHFWHQIQGDKDTTVTLENATWTMPSDTTLQN	852
35	orf1	159	LNLDNATITLNSAYRHDAAGAQTGSATDAPXXXXXXXXLLXVTPPTSVESRFNTLTVN L L+N+T+TLNSAY + S+ +AP L T PTS E RFNTLTVN	218
	hap	853	L L+N+T+TLNSAY + S+ +AP L T PTS E RFNTLTVN LTLNNSTVTLNSAYSASSNNAPRHRRSLETETTPTSAEHRFNTLTVN	899
	orf1	219	GKLNGQGTFRFMSELFGYRSDKLKLAESSEGTYTLAVNNTGNEPASLEQLTVVEGKDNKP : GKL+GQGTF+F S LFGY+SDKLKL+ +EG YTL+V NTG EP +LEQLT++E DNKP	278
40	hap	900	GKLSGQGTFQFTSSLFGYKSDKLKLSNDAEGDYTLSVRNTGKEPVTLEQLTLIESLDNKP	959
	orf1	279	LSENLNFTLQNEHVDAGA 296 LS+ L FTL+N+HVDAGA	
	hap		LSDKLKFTLENDHVDAGA 977	
45	Amino acids	1192	2-1450 of ORF1 show 41% aa identity with hap protein in 259aa over	lap:
	Orf1	1	LDRVFAEDRRNAVWTSGIRDTKHYRSQDFRAYRQQTDLRQIGMQKNLGSGRVGILFSHNR	60
	hap	1135	LDR+F + ++AVWT+ +D + Y S FRAY+Q+T+LRQIG+QK L +GR+G +FSH+R LDRLFVDQAQSAVWTNIAQDKRRYDSDAFRAYQQKTNLRQIGVQKALANGRIGAVFSHSR	119
50	orfl	61	TENTFDDGIGNSARLAHGAVFGQYGIDRFYXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	120
	hap	1195	++NTFD+ + N A L + F QY K R+ ++YG SDNTFDEQVKNHATLTMMSGFAQYQWGDLQFGVNVGTGISASKMAEEQSRKIHRKAINYG	125
55	orf1	121	IQARYRAGFGGFGIEPHIGATRYFVQKADYRYENVNIATPGLAFNRYRAGIKADYSFKPA + A Y+ G GI+P+ G RYF+++ +Y+ E V + TP LAFNPY ACI+ DY+E R	180
55	hap	1255	+ A Y+ G GI+P+ G RYF+++ +Y+ E V + TP LAFNRY AGI+ DY+F P VNASYQFRLGQLGIQPYFGVNRYFIERENYQSEEVRVKTPSLAFNRYNAGIRVDYTFTPT	131
	orf1	181	QHISITPYLSLSYTDAASGKVRTRVNTAVLAQDFGKTRSAEWGVNAEIKGFTLSLHAAAA	240
60	hap	1315	+IS+ PY ++Y D ++ V+T VN VL Q FG+ E G+ AEI F +S + + DNISVKPYFFVNYVDVSNANVQTTVNLTVLQQPFGRYWQKEVGLKAEILHFQISAFISKS	137
	orfl	241	KGPQLEAQHSAGIKLGYRW 259	
65	hap	1375	+G QL Q + G+KLGYRW QGSQLGKQQNVGVKLGYRW 1393	
UJ				

WO 99/24578 PCT/IB98/01665

## Homology with a predicted ORF from N.gonorrhoeae

The blocks of ORF1 show 83.5%, 88.3%, and 97.7% identities in 467, 298, and 259 as overlap, respectively with a predicted ORF (ORF1ng) from *N.gonorrhoeae*:

5	orfl.pep	MKTTDKRTTETHRKAPKTGRIRFXAAYLAICLSFGILPQAWAGHTYFGINYQYYRDFAEN	60
	orflng	MKTTDKRTTETHRKAPKTGRIRFSPAYLAICLSFGILPQARAGHTYFGINYQYYRDFAEN	60
••	orf1.pep	GKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVVSRNGVAALVGVQYIVSVAHNGGYN 	120
10	orflng	KGKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVVSRNGVAALAGDQYIVSVAHNGGYN	120
	orfl.pep	NVDFGAEGXNIXDQXRXTYKIVKRNNYKAGTKGHPYGGDYHMPRLHKXVTDAEPVEMTSY	180
15	orflng	NVDFGAEGSN-PDQHRFSYQIVKRNNYKAGTNGHPYGGDYHMPRLHKFVTDAEPVEMTSY	179
	orfl.pep	MDGRKYIDQNNYPDRVRIGAGRQYWRSDEDEPNNRESSYHIAS	223
	orflng	MDGWKYADLNKYPDRVRIGAGRQYWRSDEDEPNNRESSYHIASAYSWLVGGNTFAQNGSG	239
20	orfl.pep	GSPMFIYDAQKQKWLIN <u>GVLOTGNPYIGKSNG</u>	255
	orflng	GGTVNLGSEKIKHSPYGFLPTGGSFGDSGSPMFIYDAQKQKWLINGVLQTGNPYIGKSNG	289
25	orf1.pep	FOLVRKDWFYDEIFAGDTHSVFYEPRQNGKYSFNDDNNGTGKINAKHEHNSLPNRLKTRT	315
23	orflng	<u>FOL</u> VRKDWFYDEIFAGDTHSVFYEPHQNGKYFFNDNNNGAGKIDAKHKHYSLPYRLKTRT	359
	orfl.pep	VQLFNVSLSETAREPVYHAAGGVNSYRPRLNNGENISFIDEGKGELILTSNINQGAGGLY	375
30	orflng	VQLFNVSLSETAREPVYHAAGGVNSYRPRLNNGENISFIDKGKGELILTSNINQGAGGLY	
	orfl.pep	FQGDFTVSPENNETWQGAGVHISEDSTVTWKVNGVANDRLSKIGKGT	422
35	orflng	FEGNFTVSPKNNETWQGAGVHISDGSTVTWKVNGVANDRLSKIGKGTLLVQAKGENQGSV	479
33	orf1.pep	DKVTASLTKTDISGNVDLADHAHLNLTGLA	744
	orflng		774
40	orfl.pep	TLNGNLSANGDTR-YTVSHNATQNGNXSLVXNAQATFNQATLNGNTSASGNASFNLSDHA	803
	orf1ng	:     ::::   :	833
45	orfl.pep	VQNGSLTLSGNAKANVSHSALNGNVSLADKAVFHFESSRFTGQISGGKDTALHLKDSEWT	863
40	orflng	VQNGSLTLSDNAKANVSHSALNGNVSLADKAVFHFENSRFTGKISGGKDTALHLKDSEWT	893
	orfl.pep	${\tt LPSGXELGNLNLDNATITLNSAYRHDAAGAQTGSATDAPRRSRRSRRSLLXVTPPTSVE}$	923
50	orflng	:	950
	orfl.pep	SRFNTLTVNGKLNGQGTFRFMSELFGYRSDKLKLAESSEGTYTLAVNNTGNEPASLEQLT	983
<i>E.E.</i>	orflng		1010
55	orfl.pep	VVEGKDNKPLSENLNFTLQNEHVDAGAW	1011
	orflng		1070
60	orfl.pep	// LDRVFAEDRRNAVWTSGIRDTKHYRSQDFR	1211
	orf1ng		1239
6.5	orfl.pep	AYRQQTDLRQIGMQKNLGSGRVGILFSHNRTENTFDDGIGNSARLAHGAVFGQYGIDRFY	1271
65	orf1ng		1299
	-		

	orfl.pep	IGISAGAGFSSGSLSDGIGXKXRRRVLHYGIQARYRAGFGGFGIEPHIGATRYFVQKADY 1331
5	•	RYENVNIATPGLAFNRYRAGIKADYSFKPAQHISITPYLSLSYTDAASGKVRTRVNTAVL 1391
10	•	AQDFGKTRSAEWGVNAEIKGFTLSLHAAAAKGPQLEAQHSAGIKLGYRW 1440 

# The complete length ORF1ng nucleotide sequence was identified <SEQ ID 653>:

	1	ATGAAAACAA	CCGACAAACG	GACAACCGAA	ACACACCCCA	AACCCCCTAA
	51	AACCGGCCGC	ATCCGCTTCT	CGCCCGCTTA	CTTAGCCATA	でにしている。
15	101	TCGGCATTCT	GCCCCAAGCC	CGGGCGGGAC	ACACTTATTT	CGGCATCAAC
	151	TACCAATACT	ATCGCGACTT	TGCCGAAAAT	AAAGGCAAGT	TTGCAGTCGG
	201	GGCGAAAGAT	ATTGAGGTTT	ACAACAAAAA	AGGGGAGTTG	GTCGGCAAAT
÷	251	CGATGACGAA	AGCCCCGATG	ATTGATTTTT	CTGTGGTATC	GCGTAACGGC
	301	GTGGCGGCAT	TGGCGGGCGA	TCAATATATT	GTGAGCGTGG	CACATAACGC
20	351	CGGCTATAAC	AATGTTGATT	TTGGTGCGGA	GGGAAGCAAT	CCCGATCAGC
	401	ACCGCTTTTC	TTACCAAATT	GTGAAAAGAA	ATAATTATAA	AGCAGGGACT
	451	AACGGCCATC	CTTATGGCGG	CGATTATCAT	ATGCCGCGTT	TGCACAAATT
	501	TGTCACAGAT	GCAGAACCTG	TTGAGATGAC	CAGTTATATG	GATGGGTGGA
11	551	AATACGCTGA	TTTAAATAAA	TACCCTGATC	GTGTTCGAAT	CGGAGCAGGC
25	601	AGAÇAATATT	GGCGGTCTGA	TGAAGACGAA	CCCAATAACC	GCGAAAGTTC
	651	ATATCATATT	GCAAGCGCAT	ATTCTTGGCT	CGTCGGTGGC	AATACCTTTC
	701	CACAAAATGG	ATCAGGTGGT	GGCACAGTCA	ACTTAGGTAG	CGAAAAAATT
	751	AAACATAGCC	CATATGGTTT	TTTACCAACA	GGAGGCTCAT	TTGGCGACAG
	801	TGGCTCACCA	ATGTTTATCT	ATGATGCCCA	AAAGCAAAAG	ፐርርጥጥልልጥጥል
30	851	ATGGGGTATT	GCAAACAGGC	AACCCCTATA	TAGGAAAAAG	CAATGGCTTC
	901	CAGCTAGTTC	GTAAAGATTG	GTTCTATGAT	GAAATCTTTG	CTGGAGATAC
	951	CCATTCAGTA	TTCTACGAAC	CACATCAAAA	TGGGAAATAC	ተተጥጥጥጥል ልርር
•	1001	ACAATAATAA	TGGCGCAGGA	AAAATCGATG	CCAAACATAA	ACACTATTCT
	1051	CTACCTTATA	GATTAAAAAC	ACGAACCGTT	CAATTGTTTA	ATGTTTCTTT
35	1101	ATCCGAGACA	GCAAGAGAAC	CTGTTTATCA.	TGCTGCAGGT	GGGGTCAACA
	1151	GTTATCGACC	CAGACTGAAT	AATGGAGAAA	ATATTTCCTT	ТАТТСАСААА
	1201	GGAAAAGGTG	AATTGATACT	TACCAGCAAC	ATCAACCAAG	GCGCGGGCGG
	1251	TTTGTATTT	GAGGGTAATT	TTACGGTCTC	GCCTAAAAAC	AACGAAACGT
	1301	GGCAAGGCGC	GGGCGTTCAT	ATCAGTGATG	GCAGTACCGT	TACTTGGAAA
40	1351	GTAAACGGCG	TGGCAAACGA	CCGCCTGTCC	AAAATCGGCA	AAGGCACGCT
	1401	GCTGGTTCAA	GCCAAAGGGG	AAAACCAAGG	CTCGGTCAGC	GTGGGCGACG
	1451	GTAAAGTCAT	CTTAGATCAG	CAGGCGGACG	ATCAAGGCAA	AAAACAAGCC
	1501	TTTAGTGAAA	TCGGCTTGGT	CAGCGGCAGG	GGGACGGTGC	AACTGAATGC .
45	1551	CGATAATÇAG	TTCAACCCCG	ACAAACTCTA	TTTCGGCTTT	CGCGGCGGAC
45	1601	GTTTGGATTT	GAACGGGCAT	TCGCTTTCGT	TCCACCGCAT	ТСАВВАТАСС
	1651	GATGAAGGGG	CGATGATTGT	CAACCACAAT	CAAGACAAAG	AATCCACCGT
	1701	TACCATTACA	GGCAATAAAG	ATATTACTAC	AACCGGCAAT	AACAACAACT
•	1751	TGGATAGCAA	AAAAGAAATT	GCCTACAACG	GTTGGTTTGG	CGAGAAAGAT
	1801	GCAACCAAAA	CGAACGGGCG	GCTCAATCTG	AATTACCAAC	CGGAAGAAGC
50	1851	GGATCGCACT	TTACTGCTTT	CCGGCGGAAC	AAATTTAAAC	GGCAATATCA
	1901	CGCAAACAAA	CGGCAAACTG	TTTTTCAGCG	GCAGACCGAC	ACCGCACGCC
	1951	TACAATCATT	TAGGAAGCGG	GTGGTCAAAA	ATGGAAGGTA	TCCCACAAGG
	2001	AGAAATCGTG	TGGGACAACG	ATTGGATCGA	CCGCACATTT	AAAGCGGAAA
c c	2051	ACTTCCATAT	TCAGGGCGGA	CAAGCGGTGG	TTTCCCGCAA	TGTTGCCAAA
55	2101	GTGGAAGGCG	ATTGGCATTT	AAGCAATCAC	GCCCAAGCAG	TTTTCGGTGT
•	2151	CGCACCGCAT	CAAAGCCACA	CAATCTGTAC	ACGTTCGGAC	TGGACGCGTC
•	2201	TGACAAGTTG	TACCGAAAAA	ACCATTACCG	ACGATAAAGT	GATTGCTTCA
•	2251	TTGAGCAAGA	CCGACATCAG	AGGCAATGTC	AGCCTTGCCG	ATCACGCTCA
<b>60</b>	2301	TTTAAATCTC	ACAGGACTTG	CCACACTCAA	CGGCAATCTT	AGTGCAGGCG
60	2351	GAGACACGCA	CTATACGGTT	ACGCGCAACG	CCACCCAAAA	CGGCAACCTC
	2401	AGCCTCGTGG	GCAATGCCCA	AGCAACATTT	AATCAAGCCA	CATTAAACGG
	2451	CAACACATCG	GCTTCGGACA	ATGCTTCATT	TAATCTAAGC	AACAACGCCG
	2501	TACAAAACGG	CAGTCTGACG	CTTTCCGACA	ACGCTAAGGC	AAACGTAAGC
65	2551	CATTCCGCAC	TCAACGGCAA	TGTCTCCCTA	GCCGATAAGG	CAGTATTCCA
65	2601	TTTTGAAAAC	AGCCGCTTTA	CCGGAAAAAT	CAGCGGCGGC	AAGGATACGG
	2651	CATTACACTT	AAAAGACAGC	GAATGGACGC	TGCCGTCGGG	CACGGAATTA
	2701	GGCAATTTAA	ACCTTGACAA	CGCCACCATT	ACACTCAATT	CCGCCTATCG
•	2751	ACACGATGCG	GCAGGCGCGC	AAACCGGCAG	TGCGGCAGAT	GCGCCGCGCC
<b>50</b>	2801	GCCGTTCGCG	CCGTTCCCTA	TTATCCGTTA	CGCCGCCAAC	TTCGGCAGAA
70	2851	TCCCGTTTCA	ACACGCTGAC	GGTAAACGGC	AAATTGAACG	GTCAGGGAAC
·						

	2901	ATTCCGCTTT	ATGTCGGAAC	TCTTCGGCTA	CCGCAGCGGC	AAATTGAAGC
	2951	TGGCGGAAAG	TTCCGAAGGC	ACTTACACCT	TGGCTGTCAA	CAATACCGGC
	3001	AACGAACCCG	TAAGTCTCGA	GCAATTGACG	GTAGTGGAAG	GAAAAGACAA
	3051	CACACCGCTG	TCCGAAAATC	TTAATTTCAC	CCTGCaaaAc	gaacacgtcg
5	3101	atoccoococ	atggCGTTAT	CAGCTTATCC	gcaaagacgG	CGAGTTCCgc
3	3151	CTGCATAATC	CGGTCAAAGA	ACAAGAGCTT	TCCGACAAAC	TCGGCAAGgc
	3201	gggagaaACA	GAggccgccT	TGACGGCAAA	ACAGGCacaA	CTTGCCGCCA
	3251	AAcaacaggc	ggaaaAAGAC	AACgcgcaaa	gccttgAcgc	gctgattgcg
	3301	aCcagacaca	atqccaccqa	AAAGGCAgaa	agtgttgccg	aaccgGCCCG
10	3351	GCAGGCAGGC	GGGGAAAAtq	ccgGCATTAT	GCAGGCGGAG	GAAGAGAAAA
10	3401	AACGGGTGCA	GGCGGATAAA	GACACCGCCT	TGGCGAAACA	GCGCGAAGCG
	3451	GAAACCCGGC	CGGCTACCAC	CGCCTTCCCC	CGCGCCCGCC	GCGCCCGCCG
	3501	GGATTTGCCG	CAACCGCAGC	CCCAACCGCA	ACCCCAACCG	CAGCGCGACC
	3551	TGATCAGCCG	TTATGCCAAT	AGCGGTTTGA	GTGAATTTTC	CGCCACGCTC
15	3601	AACAGCGTTT	TCGCCGTACA	GGACGAATTG	GACCGCGTGT	TTGCCGAAGA
10	3651	CCGCCGCAAC	GCCGTTTGGA	CAAGCGGCAT	CCGGGACACC	AAACACTACC
	3701	GTTCGCAAGA	TTTCCGCGCC	TACCGCCAAC	AAACCGACCT	GCGCCAAATC
	3751	GGTATGCAGA	AAAACCTCGG	CAGCGGGCGC	GTCGGCATCC	TGTTTTCGCA
	3801	CAACCGGACC	GGAAACACCT	TCGACGACGG	CATCGGCAAC	TCGGCACGGC
20	3851	TTGCCCACGG	TGCCGTTTTC	GGGCAATACG	GCATCGGCAG	GTTCGACATC
	3901	GGCATCAGCG	CGGGCGCGGG	TTTTAGTAGC	GGCAGCCTTT	CAGACGGCAT
	3951	CAGAGGCAAA	ATCCGCCGCC	GCGTGCTGCA	TTACGGCATT	CAGGCAAGAT
	4001	ACCGCGCAGG	TTTCGGCGGA	TTCGGCATCG	AACCGCACAT	CGGCGCAACG
	4051	CGCTATTTCG	TCCAAAAAGC	GGATTACCGA	TACGAAAACG	TCAATATCGC
25	4101	CACCCCGGGC	CTTGCATTCA	ACCGCTACCG	CGCGGGCATT	AAGGCAGATT
	4151	ATTCATTCAA	ACCGGCGCAA	CACATTTCCA	TCACGCCTTA	TTTGAGCCTG
	4201	TCCTATACCG	ATGCCGCTTC	CGGCAAAGTC	CGAACGCGCG	TCAATACCGC
	4251	CGTATTGGCG	CAGGATTTCG	GCAAAACCCG	CAGTGCGGAA	TGGGGCGTAA
	4301	ACGCCGAAAT	CAAAGGTTTC	ACGCTGTCCC	TCCACGCTGC	CGCCGCCAAG
30	4351	GGGCCGCAAT	TGGAAGCGCA	GCACAGCGCG	GGCATCAAAT	TAGGCTACCG
<del>-</del> -	4401	CTGGTAA				

This is predicted to encode a protein having amino acid sequence <SEQ ID 654>:

	1	MKTTDKRTTE	THRKAPKTGR	IRFSPAYLAI	CLSFGILPQA	RAGHTYFGIN
	51	YQYYRDFAEN	KGKFAVGAKD	IEVYNKKGEL	VGKSMTKAPM	IDFSVVSRNG
35	101	VAALAGDQYI	VSVAHNGGYN	NVDFGAEGSN	PDQHRFSYQI	VKRNNYKAGT
	151	NGHPYGGDYH	MPRLHKFVTD	AEPVEMTSYM	DGWKYADLNK	YPDRVRIGAG
	201	RQYWRSDEDE	PNNRESSYHI	ASAYSWLVGG	NTFAQNGSGG	GTVNLGSEKI
	251	KHSPYGFLPT	GGSFGDSGSP	MFIYDAQKQK	WLINGVLOTG	NPYIGKSNGF
	301	OLVRKDWFYD	EIFAGDTHSV	FYEPHQNGKY	FFNDNNNGAG	KIDAKHKHYS
40	351	LPYRLKTRTV	QLFNVSLSET	AREPVYHAAG	GVNSYRPRLN	NGENISFIDK
••	401	GKGELILTSN	INQGAGGLYF	EGNFTVSPKN	NETWQGAGVH	ISDGSTVTWK
	451	VNGVANDRLS	KIGKGTLLVQ	AKGENQGSVS	VGDGKVILDQ	QADDQGKKQA
	501	FSEIGLVSGR	GTVQLNADNQ	FNPDKLYFGF	RGGRLDLNGH	SLSFHRIQNT
	551	DEGAMIVNHN	QDKESTVTIT	GNKDITTTGN	NNNLDSKKEI	AYNGWFGEKD
45	601	ATKTNGGLNL	NYPPEEADRT	LLLSGGTNLN	GNITQTNGKL	FFSGRPTPHA
	651	YNHLGSGWSK	MEGIPQGEIV	WDNDWIDRTF	KAENFHIQGG	QAVVSRNVAK
	701	VEGDWHLSNH	AQAVFGVAPH	QSHTICTRSD	WTGLTSCTEK	TITDDKVIAS
	751	LSKTDVRGNV	SLADHAHLNL	TGLATFNGNL	VQAETRTIRL	RANATQNGNL
	801	SLVGNAQATF	NQATLNGNTS	ASDNASFNLS	NNAVQNGSLT	LSDNAKANVS
50	851	HSALNGNVSL	ADKAVFHFEN	SRFTGKISGG	KDTALHLKDS	EWTLPSGTEL
	901	GNLNLDNATI	TLNSAYRHDA	AGAQTGSAAD	APRRRSRRSL	LSVTPPTSAE
	951	SRFNTLTVNG	KLNGQGTFRF	MSELFGYRSG	KLKLAESSEG	TYTLAVNNTG
	1001	NEPVSLEQLT	VVEGKDNTPL	SENLNFTLQN	EHVDAGAWRY	QLIRKDGEFR
	1051	LHNPVKEQEL	SDKLGKAGET	EAALTAKQAQ	LAAKQQAEKD	NAQSLDALIA
55	1101	AGRNATEKAE	SVAEPARQAG	GENAGIMQAE	EEKKRVQADK	DTALAKQREA
	1151	ETRPATTAFP	RARRARRDLP	QPQPQPQPQP	QRDLISRYAN	SGLSEFSATL
	1201	NSVFAVQDEL	DRVFAEDRRN	AVWTSGIRDT	KHYRSQDFRA	YRQQTDLRQI
	1251	GMQKNLGSGR	VGILFSHNRT	GNTFDDGIGN	SARLAHGAVF	GQYGIGRFDI
	1301	GISAGAGFSS	GSLSDGIRGK	IRRRVLHYGI	QARYRAGFGG	FGIEPHIGAT
60	1351	RYFVQKADYR	YENVNIATPG	LAFNRYRAGI	KADYSFKPAQ	HISITPYLSL
	1401	SYTDAASGKV	RTRVNTAVLA	QDFGKTRSAE	WGVNAEIKGF	TLSLHAAAAK
	1451	GPQLEAQHSA	GIKLGYRW*			

Underlined and double-underlined sequences represent the active site of a serine protease (trypsin family) and an ATP/GTP-binding site motif A (P-loop).

ORF1-1 and ORF1ng show 93.7% identity in 1471 aa overlap:

	orfl-1.pep	10 MKTTDKRTTET	20 HRKAPKTGRIRF	30 SPAYLAICLSE	40 GILPOAWAGH	50 TYFGINYOYY	60 RDFAEN
5	orflng-1				 GILPQARAGH	 TYFGINYQYY	111111 RDFAEN
3	•	70	. 80	30 90	40 100	50 110	60 120
	orf1-1.pep		EVYNKKGELVGK	SMTKAPMIDES	VVSRNGVAAL	VGDQYIVSVA	HNGGYN
10	orflng-1	KGKFAVGAKDI 70	EVYNKKGELVGK 80	SMTKAPMIDFS 90	VVSRNGVAAL 100	AGDOYIVSVA 110	HNGGYN 120
	orf1-1.pep	130	140 DQHRFTYKIVKR	150	160	170	180
15	orflng-1		: :     DQHRFSYQIVKR	:    NNYKAGTNGHE	 YGGDYHMPRL	 HKFVTDAEPV	11111
		130	200	150 210	160 220	170 230	180
20	orf1-1.pep	DGRKYIDQNNY	PDRVRIGAGROY	WRSDEDEPNNE	ESSYHIASAY	SWLVGGNTFA	240 AQNGSGG
	orf1ng-1	DGWKYADLNKY 190	PDRVRIGAGRQY 200	WRSDEDEPNNF 210	ESSYHIASAY 220	SWLVGGNTFA 230	AQNGSGG 240
25	orf1-1.pep	250 GTVNLGSEKIK	260 HSPYGFLPTGGS	270 FGDSGSPMFIY	280 DAQKQKWLIN	290. GVLOTGNPYI	300 GKSNGF
	orflng-1	GTVNLGSEKIK			 DAQKQKWLIN		GKSNGF
30		250 310	260 320	270	280	290	300
	orf1-1.pep	QLVRKDWFYDE	IFAGDTHSVFYE	330 PRQNGKYSFNI  :	340 DNNGTGKINA :   :   :	350 KHEHNSLPNE	360 RLKTRTV
35	orflng-1	QLVRKDWFYDE 310	IFAGDTHSVFYE 320	PHONGKYFFNE 330	NNNGAGKIDA 340	KHKHYSLPYR 350	LKTRTV 360
•	orf1-1.pep	370 QLFNVSLSETA	380 REPVYHAAGGVN	390 SYRPRLNNGEN	400 HISFIDEGKGE	410 LILTSNINQG	420 SAGGLYF
40	orflng-1	QLFNVSLSETA		IIIIIIIIIIIIIIIISYRPRLNNGEN 390	:       ISFIDKGKGE  400	LILTSNINQG	HIIIIII GAGGLYF 420
		430	440	450	460	470	480
45	orfl-1.pep	:::::::::::::::::::::::::::::::::::::::	ETWQGAGVHISE	1111111111	11111111111	111 311111	1111:1:
	orflng-1	EGNFTVSPKNN 430	ETWQGAGVHISD 440	GSTVTWKVNGV 450	ANDRLSKIGK 460	GTLLVQAKGE 470	ENQGSVS 480
50	orfl-1.pep	490 VGDGTVILDQQ	500 ADDKGKKQAFSE	510 IGLVSGRGTVC	520 LNADNOFNPD	530 KLYFGFRGGF	540 RLDLNGH
	orflng-1	VGDGKVILDQQ 490	:         ADDQGKKQAFSE 500	IGLVSGRGTVC 510		KLYFGFRGGF	
55	orf1-1.pep	550	560	570	580	590	600
	orflng-1	-	EGAMIVNHNQDK           EGAMIVNHNQDK		1:111111:1	1111111111111	
60	<b></b>	550	560	570	580	590	600
	orfl-1.pep	610 TTKTNGRLNLV	620 YQPAAEDRTLLL	630 SGGTNLNGNIT	640 OTNGKLFFSG	650 RPTPHAYNHI	660 CNDHWSQ
65	orflng-1	:	III IIIIII YQPEEADRTLLL 620	IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII			GSGWSK 660
		670	680	690	700	710	720
70	orfl-1.pep		DNDWINRTFKAE	11:1:11111	SRNVAKVKGD	WHLSNHAQAV	/FGVAPH
70	orflng-1	MEGIPQGEIVW 670	DNDWIDRTFKAE 680	NFHIQGGQAVV 690	SRNVAKVEGD 700	WHLSNHAQAV 710	FGVAPH 720

	orf1-1.pep	730 740 750 760 770 780 QSHTICTRSDWTGLTNCVEKTITDDKVIASLTKTDISGNVDLADHAHLNLTGLATLNGNL
5	orf1ng-1	
	orf1-1.pep	790 800 810 820 830 840 SANGDTRYTVSHNATQNGNLSLVGNAQATFNQATLNGNTSASGNASFNLSDHAVQNGSLT
10	orflng-1	SAGGDTHYTVTRNATQNGNLSLVGNAQATFNQATLNGNTSASDNASFNLSNNAVQNGSLT 790 800 810 820 830 840
15	orfl-1.pep	850 860 870 880 890 900 LSGNAKANVSHSALNGNVSLADKAVFHFESSRFTGQISGGKDTALHLKDSEWTLPSGTEL
	orflng-1	LSDNAKANVSHSALNGNVSLADKAVFHFENSRFTGKISGGKDTALHLKDSEWTLPSGTEL 850 860 870 880 890 900
20	orfl-1.pep	910 920 930 940 950 960 GNLNLDNATITLNSAYRHDAAGAQTGSATDAPRRSRRSRRSLLSVTPPTSVESRFNTLT
	orflng-1	
25	orf1-1.pep	970 980 990 1000 1010 1020 VNGKLNGQGTFRFMSELFGYRSDKLKLAESSEGTYTLAVNNTGNEPASLEQLTVVEGKDN
	orflng-1	
30		960 970 980 990 1000 1010 1030 1040 1050 1060 1070
	orf1-1.pep	KPLSENLNFTLQNEHVDAGAWRYQLIRKDGEFRLHNPVKEQELSDKLGKA
35	orflng-1	TPLSENLNFTLQNEHVDAGAWRYQLIRKDGEFRLHNPVKEQELSDKLGKAGETEAALTAK 1020 1030 1040 1050 1060 1070
	orfl-1.pep	1080 1090 1100 1110 1120EAKKQAEKDNAQSLDALIAAGRDAVEKTESVAEPARQAGGENVGIMQAEEEKKRVQ
40	orflng-1	:
15	orf1-1.pep	1130 1140 1150 1160 1170 1180 ADKDTALAKQREAETRPATTAFPRARRARRDLPQLQPQPQPQPQRDLISRYANSGLSEFS
45	orflng-1	
50	orf1-1.pep	1190 1200 1210 1220 1230 1240 ATLNSVFAVQDELDRVFAEDRRNAVWTSGIRDTKHYRSQDFRAYRQQTDLRQIGMQKNLG
	orflng-1	
55	orf1-1.pep	1250 1260 1270 1280 1290 1300 SGRVGILFSHNRTENTFDDGIGNSARLAHGAVFGQYGIDRFYIGISAGAGFSSGSLSDGI
	orflng-1	SGRVGILFSHNRTGNTFDDGIGNSARLAHGAVFGQYGIGRFDIGISAGAGFSSGSLSDGI
60	·	1260 1270 1280 1290 1300 1310 1310 1320 1330 1340 1350 1360
60	orf1-1.pep	1260 1270 1280 1290 1300 1310  1310 1320 1330 1340 1350 1360  GGKIRRVLHYGIQARYRAGFGGFGIEPHIGATRYFVQKADYRYENVNIATPGLAFNRYR
60	·	1260 1270 1280 1290 1300 1310  1310 1320 1330 1340 1350 1360  GGKIRRVLHYGIQARYRAGFGGFGIEPHIGATRYFVQKADYRYENVNIATPGLAFNRYR
	orf1-1.pep	1260 1270 1280 1290 1300 1310  1310 1320 1330 1340 1350 1360  GGKIRRVLHYGIQARYRAGFGGFGIEPHIGATRYFVQKADYRYENVNIATPGLAFNRYR

		1430	1440	1450	
	orf1-1.pep	KGFTLS	LHAAAAKGPQI	EAQHSAGI	<b>LGYRWX</b>
		111111	111111111111	11111111	
	orflng-1	KGFTLS	LHAAAAKGPQI	<i>E</i> AQHSAGII	CLGYRWX
5		1440	1450	1460	

## In addition, ORF1ng shows 55.7% identity with hap protein (P45387) over a 1455aa overlap:

	In addition, ORF ing	snows 55.7% identity with nap protein (P45387) over a 14	issaa overiapi
10	SCORES I Smith-Waterma	nitl: 1104 Initn: 4632 Opt: 2680 n score: 5165; 55.7% identity in 1455 aa overlap	
10	orflng-1.pep	10 20 30 40 50 MKTTDKRTTETHRKAPKTGRIRFSPAYLAICLSFGILPQARAGHTYFGIN	
15	p45387	: : : :  :          : MKKTVFRLNFLTACISLGIVSQAWAGHTYFGID 10 20 30	
	orflng-1.pep	70 80 90 100 110 KGKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVVSRNGVAALAGDQYI	VSVAHNGGYN
20	p45387	:  :: :   : :	
	orflng-1.pep	130 140 150 160 170 NVDFGAEGSNPDQHRFSYQIVKRNNYKAGTNGHPYGGDYHMPRLHKFVTD	AE PVEMTSYM
25	p45387	:       :      :  :	
30	orflng-1.pep	190 200 210 220 230 DGWKYADLNKYPDRVRIGAGRQYWRSDEDEPNNRESSYHIASAYSWLVGG	NT FAQNGSGG
	p45387	:   : :  :   :   :  :  :  : :: :: ::  :	: : NTHNQRGAGN 210
35	orflng-1.pep	250 260 270 280 290 GTVNLGSEKIKHSPYGFLPTGGSFGDSGSPMFIYDAQKQKWLINGVLQTG	NPYIGKSNGF
40	p45387	GYSYLGGDVRKAGEYGPLPIAGSKGDSGSPMFIYDAEKQKWLINGILREG 220 230 240 250 260	270
	orflng-1.pep	310 320 330 340 350 QLVRKDWFYDEIFAGDTHSVFYEPHQNGKYFFNDNNNGAGKIDAKHKHYS	LPYRLKTRTV
45	p45387	QLVRKSYF-DEIFERDLHTSLYTRAGNGVYTISGNDNGQGSITQKSG 280 290 300 310 320	;IPSEIKI
•	orflng-1.pep	370 380 390 400 41  QLFNVSLSETARE PVYHAA-GGVNSYRPRLNNGENISFIDKGKGELILTS     :     :   :                     :   :   :     :       :	NINQGAGGLY
50	p45387	TLANMSLPLKEKDKVHNPRYDGPNIYSPRLNNGETLYFMDOKOGSLIFAS 330 340 350 360 370 38	DINQGAGGLY
55	orf1ng-1.pep p45387		/QAKGENQGSV       :  : /QAKGENKGSI
		480 490 500 510 520 53	10 30 539
60	orf1ng-1.pep p45387	SVGDGKVILDQQADDQGKKQAFSEIGLVSGRGTVQLNADNQFNPDKLYFC	1111111111
65	orflng-1.pep p45387	540 550 560 570 580 HSLSFHRIQNTDEGAMIVNHNQDKESTVTITGNKDITT-TGNN-NNLDSI	590 KKEIAYNGWFG
70	passor		60

5	orflng-1.pep p45387	600 610 620 630 640 650 EKDATKTNGRLNLNYQPEEADRTLLLSGGTNLNGNITQTNGKLFFSGRPTPHA     :	YNHLGSG
10	orflng-1.pep p45387	660 670 680 690 700 710 WSKMEGIPQGEIVWDNDWIDRTFKAENFHIQGGQAVVSRNVAKVEGDWHLSNH   :          : :  :  :  :  :  :  :  :	AQAVFGV
15	orflng-1.pep	720 730 740 750 760 770  APHQSHTICTRSDWTGLTSCTEKTITDDKVIASLSKTDIRGNVSLADHAHLNI : :::            : ::	LTGLATIN
20	orflng-1.pep	780 790 800 810 820 830 GNLSAGGDTHYTVTRNATQNGNLSLVGNAQATFNQATLNGNTSASDNASFNL   :: :::: :  GNVTLTNHSQFTL 750 760	SNNAVQNG
25	orflng-l.pep	840 850 860 870 880 89 SLTLSDNAKANVSHSALNGNVSLADKAVFHFENSRFTGKISGGKDTALHLKD ::     :  : :       : :   :: !: :  :: : : :: :: :: :: :: :: :: :: :: :	SEWTLPSG:   :
30	p45387	NIRLSDNSTATVDNANLNGNVHLTDSAQFSLKNSHFSHQIQGDKGTTVTLEN         780       790       800       810       820         900       910       920       930       940       95	830
35	orflng-1.pep p45387	TELGNLNLDNATITLNSAYRHDAAGAQTGSAADAPRRSRRSLLSVTPPTSA	1 111111
40	orflng-1.pep	960 970 980 990 1000 101 VNGKLNGQGTFRFMSELFGYRSGKLKLAESSEGTYTLAVNNTGNEPVSLEQI      :    :	TVVEGKDN
45	orflng-1.pep	1020 1030 1040 1050 1060 107  TPLSENLNFTLQNEHVDAGAWRYQLIRKDGEFRLHNPVKEQELSDKLGKAGE    :: :  :  :          : :   QPLSDKLKFTLENDHVDAGALRYKLVKNDGEFRLHNPIKEQELHNDLVRAEQ 940 950 960 970 980 990	ETEAALTAK
50	orflng-1.per	1080 1090 1100 1110 1120 13  QAQLAAKQOAEKDNAQSLDALIAAGRNAT-EKAESVAEPARQAGGENAGIM(  :::    ::::   : ::::   :  : ::  QVEPTAKTQTGEPKVRSRRAARAAFPDTLPDQSLLNALEAKQAE-LTAETQI	::: :  :  KSKAKTKKV
55	orflng-1.pep		1190 LISRYANSG
60	p45387	RSKRAVFSDPLLDQSLFALEAALEVIDAPQQSEKDRLAQEEAEKQ-RKQKD 1060 1070 1080 1090 1100 11	LISRYSNSA
65	orflng-1.pe	LSEFSATLNSVFAVQDELDRVFAEDRRNAVWTSGIRDTKHYRSQDFRAYRQ	Q-TDLRQIG
70	orflng-1.pe	:    ::  :  :  : :     : :        : :  :      ::: : VOKALANGRIGAVFSHSRSDNTFDEQVKNHATLTMMSGFAQYQWGDLQFGV	:: : : ::

-376-

	orflng-1.pep	::::	GKIRRRV	::  ::   :	:1:11:1	:: ::   ::	1360 OKADYRYENVN :::: : :  ERENYQSEEVR	: 11:1
5		1240	1250	1260	1270	1280	1290	
		. 130		1390	1400	1410	1420	1430
	orflng-1.pep	AFNRYRA	SIKADYS	KPAQHISIT	PYLSLSYTDA	ASGKVRTRVI	NTAVLAQDEGR	TRSAEW
10	p45387	AFNRYNA( 1300	::  :  GIRVDYT    1310	:::  : FTPTDNISVK 1320	: :: : :  PYFFVNYVD\   1330	:::: :    SNANVQTTVI 1340	:     : NLTVLQQPFGR 1350	YWQKEV
		2000	1010	1320	1330	1340	1350	
		14		1450	1460	1469		
15	orflng-1.pep			AAAAKGPOLE.				
13	p45387			: ::     [SKSQGSQLG	::: :     KQQNVGVKL		1	
		1360	1370	1380	1390			

Based on this analysis, it is predicted that these proteins from N.meningitidis and N.gonorrhoeae, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### 20 **Example 78**

The following partial DNA sequence was identified in N.meningitidis <SEQ ID 655>:

```
.. AAGGTGTGGC AATTTGTCGA AGA.CCGCTG CGTGCCGTCG TGCCTGCCGA
                 51
                       CAGTTTTGAA CCGACCGCGC AAAAATTGAA CCTGTTTAAG GCGGGTGCGG
                       CAACCATTTT GTTTTATGAA GATCAAAATG TCGTCAAAGG TTTGCAGGAG
                101
25
                151
                       CAGTTCCCTG CTTATGCCGC TAACTTCCCC GTTTGGGCGG ATCAGGCAAA
                       CGCGATGGTG CAGTATGCCG TTTGGACGAC ACTTGCCGCG GTCGGCGTAG
                201
                251
                       GTGCAAACCT GCAACATTAC AATCCCTTGC CCGATGCGGC GATTGCCAAA
                       GCGTGGAATA TCCCCGAAAA CTGGTTGTTG CGCGCACAAA TGGTTATCGG
                301
                351
                       CGGTATTGAA GGGGCGCAG GTGAAAAGAC CTTTGAACCC GTTGCAGAAC
30
                401
                       GTTTGAAAGT GTTCGGCGCA TAA
```

This corresponds to the amino acid sequence <SEQ ID 656; ORF6>:

```
1 KVWQFVEXPL RAVVPADSFE PTAQKLNLFK AGAATILFYE DONVVKGLQE
51 QFPAYAANFP VWADQANAMV QYAVWTTLAA VGVGANLQHY NPLPDAAIAK
101 AWNIPENWLL RAQMVIGGIE GAAGEKTFEP VAERLKVFGA *
```

35 Further sequence analysis revealed a further partial DNA sequence <SEQ ID 657>:

```
..CTGCGTGCCG TCGTGCCTGC CGACAGTTTT GAACCGACCG CGCAAAAATT
                51
                      GAACCTGTTT AAGGCGGGTG CGGCAACCAT TTTGTTTTAT GAAGATCAAA
               101
                      ATGTCGTCAA AGGTTTGCAG GAGCAGTTCC CTGCTTATGC CGCTAACTTC
                      CCCGTTTGGG CGGATCAGGC AAACGCGATG GTGCAGTATG CCGTTTGGAC
               151
40
               201
                      GACACTTGCC GCGGTCGGCG TAGGTGCAAA CCTGCAACAT TACAATCCCT
                      TGCCCGATGC GGCGATTGCC AAAGCGTGGA ATATCCCCGA AAACTGGTTG
               251
                301
                      TTGCGCGCAC AAATGGTTAT CGGCGGTATT GAAGGGGCGG CAGGTGAAAA
                      GACCTTTGAA CCCGTTGCAG AACGTTTGAA AGTGTTCGGC GCATAA
               351
```

This corresponds to the amino acid sequence <SEQ ID 658; ORF6-1>:

```
1 ..LRAVVPADSF EPTAQKLNLF KAGAATILFY EDQNVVKGLQ EQFPAYAANF
51 PVWADQANAM VQYAVWTTLA AVGVGANLQH YNPLPDAAIA KAWNIPENWL
101 LRAQMVIGGI EGAAGEKTFE PVAERLKVFG A*
```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from N.meningitidis (strain A)

ORF6 shows 98.6% identity over a 140aa overlap with an ORF (ORF6a) from strain A of N. meningitidis:

-377-

	orf6.pep	KVWQFVEXPLRAVVPADSFEPTAQKLNLFK
	orf6a	
5		40 30 00 70 00
	orf6.pep	AGAATILFYEDONVVKGLQEQFPAYAANFPVWADQANAMVQYAVWTTLAAVGVGANLQHY
	orf6a	
10		100 110 120 130 140 150
	orf6.pep	100 110 120 130 140 NPLPDAAIAKAWNIPENWLLRAQMVIGGIEGAAGEKTFEPVAERLKVFGAX
15	orf6a	
13	Ollea	160 170 180 190 200
	The complete length	gth ORF6a nucleotide sequence <seq 659="" id=""> is:</seq>
*	1 7	ATGACCCGTC AATCTCTGCA ACAGGCTGCC GAAAGCCGCC GTTCCATTTA
20	101	TTCGTTAAAT AAAAATCTGC CCGTCGGCAA AGATGAAATC GTCCAAATCG TCGAACACGC CGTTTTGCAC ACACCTTCTT CGTTCAATTC CCAATCTGCC
	151	CGTGTGGTCG TGCTGTTTGG CGAAGAGCAT GATAAGGTGT GGCAATTTGT CGAAGACGCG CTGCGTGCCG TCGTGCCTGC CGACAGTTTT GAACCGACCG
	251	CGCAAAAATT GAACCTGTTT AAGGCGGGTG CGGCAACTAT TTTGTTTTAT
	301	GAAGATCAAA ATGTCGTCAA AGGTTTGCAG GAGCAGTTCC CTGCTTATGC
25	351 401	CGCCAACTTT CCCGTTTGGG CGGACCAGGC GAACGCGATG GTGCAGTATG CCGTTTGGAC GACACTTGCC GCGGTCGGCG TAGGTGCAAA CCTGCAACAT
	451	TACAATCCCT TGCCCGATGC GGCGATTGCC AAAGCGTGGA ATATCCCCGA
	501	AAACTGGTTG TTGCGCGCAC AAATGGTTAT CGGCGGTATT GAAGGGGCGG
30		CAGGTGAAAA GACCTTTGAA CCAGTTGCAG AACGTTTGAA AGTGTTCGGC GCATAA
	This is predicted	to encode a protein having amino acid sequence <seq 660="" id="">:</seq>
	1	MTRQSLQQAA ESRRSIYSLN KNLPVGKDEI VQIVEHAVLH TPSSFNSQSA
	51	RVVVLFGEEH DKVWQFVEDA LRAVVPADSF EPTAQKLNLF KAGAATILFY
	101	EDONVVKGLQ EQFPAYAANF PVWADQANAM VQYAVWTTLA AVGVGANLQH
35		YNPLPDAAIA KAWNIPENWL LRAQMVIGGI EGAAGEKTFE PVAERLKVFG A*
	ODDC 10DD	26.1 strong 100.00/ identity in 121 on exertent
	ORFoa and ORF	6-1 show 100.0% identity in 131 aa overlap:
		50 60 70 80 90 100
40	orf6a.pep	TPSSFNSQSARVVVLFGEEHDKVWQFVEDALRAVVPADSFEPTAQKLNLFKAGAATILFY
	orf6-1	LRAVVPADSFEPTAQKLNLFKAGAATILFY 10 20 30
45		110 120 130 140 150 160
73	orf6a.pep	EDONYVKGLOEOFPAYAANFPVWADOANAMVQYAVWTTLAAVGVGANLQHYNPLPDAAIA
	orf6-1	
	0110 1	40 50 60 70 80 90
50		170 180 190 200
	orf6a.pep	KAWNIPENWLLRAQMVIGGIEGAAGEKTFEPVAERLKVFGAX
	• •	[ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ]
55	orf6-1	KAWNIPENWLLRAQMVIGGIEGAAGEKTFEPVAERLKVFGAX 100 110 120 130
JJ		700 770 700

# Homology with a predicted ORF from N.gonorrhoeae

ORF6 shows 95.7% identity over a 140aa overlap with a predicted ORF (ORF6ng) from N.gonorrhoeae:

	orf6.pep	KVWQFVEXPLRAVVPADSFEPTAQKLNLFK	30
	orf6ng		64
5	orf6.pep	AGAATILFYEDQNVVKGLQEQFPAYAANFPVWADQANAMVQYAVWTTLAAVGVGANLQHY	90
	orf6ng	AGAATILFYEDQNVVKGLQEQFPAYAANFPVWADQANAMVQYAVWTTLAAVGAGANLQHY	124
10	orf6.pep	NPLPDAAIAKAWNIPENWLLRAQMVIGGIEGAAGEKTFEPVAERLKVFGA 140	
	orf6ng	:	

The complete length ORF6ng nucleotide sequence <SEQ ID 661> was identified as:

	1	ATGGCCGTTG	CGTCAAATGT	CAGCTTGGAT	ATGTCCAATC	CTACGGTGTT
15	51	ACGCATGGGA	TTACCCTTAT	ATATTGCGTC	CCTAAGAAGG	GGCGCAATAT
	101	ATAAGGTGTG	GCAATTTGTC	GAAGACGCGC	TGCGTGCCGT	CGTGCCTGCC
	151	GACAGTTTTG	AACCGACCGC	GCAAAAATTG	<b>AAGCTGTTTA</b>	AGGCGGGCGC
	201	GGCAACCATT	TTGTTTTATG	AAGATCAAAA	TGTCGTCAAA	GGTTTGCAGG
•	251	AGCAGTTCCC	TGCTTATGCC	GCCAACTTTC	CCGTTTGGGC	GGACCAGGCG
20	3.01	AACGCTATGG	TACAGTATGC	CGTCTGGACG	ACACTTGCCG	CGGTCGGTGC
	351	AGGTGCAAAT	CTGCAACATT	ACAACCCCTT	GCCCGATGTG	GCGATTGCTA
	401	AAGCGTGGAA	TATTCCCGAA	AACTGGCTGT	TGCGCGCGCA	AATGGTTATC
	451	GGTGGTATTG	AAGGGGcggc	aggtgaaaaa	gtctttgaac	CCGTTGCgga '
*	501	acgtttgAAA	GTGTTCGGCG	CATAA		

This encodes a protein having amino acid sequence <SEQ ID 662>: 25

```
MAVASNVSLD MSNPTVLRMG LPLYIASLRR GAIYKVWQFV EDALRAVVPA
```

DSFEPTAQKL KLFKAGAATI LFYEDQNVVK GLQEQFPAYA ANFPVWADQA

NAMVQYAVWT TLAAVGAGAN LQHYNPLPDV AIAKAWNIPE NWLLRAQMVI GGIEGAAGEK VFEPVAERLK VFGA\*

ORF6ng and ORF6-1 show 96.9% identity in 131 aa overlap:

					10	20	30
	orf6-1.pep			LRA	AVVPADSFEPT		SAATILFY
35	orf6ng	PTVLRMGLPLYI	ASLRRGAIYK	III VWQFVEDALRA		:     AOKLKLFKA0	GAATILFY
		20	30	40	50	60	70
		40	50	60	70	80	90
40	orf6-1.pep	EDQNVVKGLQEQ	FPAYAANFPV	<b>WADQANAMVQ</b> Y	/AVWTTLAAVO	VGANLQHYNI	PLPDAAIA
40		EDOMERICI OF O					
	orf6ng	EDQNVVKGLQEQ	PAIAANEPV 90	WADQANAMVQ:	AVWITLAAVO		
		80	30	100	110	120	130
		100	110	120	130		
45	orf6-1.pep	KAWNIPENWLLR	AQMVIGGIE	AAGEKT FEPV	AERLKVFGAX		•
		11111111111111	1111111111	11111:1111			
	orf6ng	KAWNIPENWLLR	AQMVIGGIEG	AAGEKVFEPV <i>I</i>	AERLKVFGAX		
		140	150	160	170		

It is predicted that the proteins from N.meningitidis and N.gonorrhoeae, and their epitopes, could 50 be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 79

30

The following partial DNA sequence was identified in N.meningitidis <SEQ ID 663>

	1	GGCTACAACT	ACCTGTTCGC	GCGCGGCAGC	CGCATCGCCA	ACTACCAAAT
55	51	CAACGGCATC	CCCGTTGCCG	ACGCGCTGGC	CGATACGGG <u>t</u>	CAATGCCAAC
	101	ACCGCCGCCT	ATGAGCGCGT	AGAAGTCGTG	CGCGGCGTGG	CGGGGCTGCT
	151	GGACGGCACG	GGCGAGCCTT	CCGCCACCGT	CAATCTGGTG	CGCAAACGCC
	201	TGACCCGCAA	GCCATTGTTT	GAAGTCCGCG	CCGAAGCgGG	CAACCGCAAA

65

```
CATTTCGGGC TGGACGCGGA CGTATCGGGC AGCCTGAACA CCGAAG.crC
                           rCTGCGCgGC CGCCTGGTTT CCACCTTCGG ACGCGCGAC TCGTGGCGGC
GGCGCGAACG CAGCCGskAT GCCGAACTCT ACGGCATTTT GGAATACGAC
                   301
                   351
                           ATCGCACCGC AAACCCGCGT CCACGCArGC ATGGACTACC AGCAGGCGAA
                   401
                           AGAAACCGCC GACGCGCCGC TCAGCTACGC CGTGTACGAC AGCCAAGGTT
 5
                   451
                           ATGCCACCGC CTTCGGCCCG AAAGACAACC CCGCCACAAA TTGGGCGAAC
AGCCACCACC GTGCGCTCAA CCTGTTCGCC GGCATCGAAC ACCGCTTCAA
                   501
                   551
                           CCAAGACTGG AAACTCAAAG CCGAATACGA CTAC..
                   601
      This corresponds to the amino acid sequence <SEQ ID 664; ORF23>:
                         ..GYNYLFARGS RIANYQINGI PVADALADTG NANTAAYERV EVVRGVAGLL
10
                            DGTGEPSATV NLVRKRLTRK PLFEVRAEAG NRKHFGLDAD VSGSLNTEXX
                            LRGRLVSTFG RGDSWRRRER SRXAELYGIL EYDIAPQTRV HAXMDYQQAK
                   101
                            ETADAPLSYA VYDSQGYATA FGPKDNPATN WANSHHRALN LFAGIEHRFN
                   151
                            QDWKLKAEYD Y..
                   201
      Further work revealed the complete nucleotide sequence <SEQ ID 665>:
15
                     1 ATGACACGCT TCAAATATTC CCTGCTGTTT GCCGCCCTGT TGCCCGTGTA
                   151 GACGGCTACA CTGTTTCCGG CACGCACACC CCGCTCGGGC TGCCCATGAC
                   201 CCTGCGCGAA ATCCCGCAGA GCGTCAGCGT CATCACATCG CAACAAATGC
20
                   251 GCGACCAAAA CATCAAAACG CTCGACCGCG CCCTGTTGCA GGCGACCGGC
301 ACCAGCCGCC AGATTTACGG CTCCGACCGC GCGGGCTACA ACTACCTGTT
                         CGCGCGCGC AGCCGCATCG CCAACTACCA AATCAACGGC ATCCCCGTTG
                    351
                    401 CCGACGCGCT GGCCGATACG GGCAATGCCA ACACCGCCGC CTATGAGCGC
                   451 GTAGAGTCG TGCGCGGTACG GGCAGATGCCA ACACCGGCGC CTAGAGGCC
501 TTCCGCCACC GTCAATCTGG TGCGCAAACG CCTGACCCGC AAGCCATTGT
551 TTGAAGTCCG CGCCGAAGCG GGCAACCGCA AACATTTCGG GCTGGACGCG
25
                    601 GACGTATCGG GCAGCCTGAA CACCGAAGGC ACGCTGCGCG GCCGCCTGGT
                    651 TTCCACCTTC GGACGCGGCG ACTCGTGGCG GCGGCGCGAA CGCAGCCGCG
701 ATGCCGAACT CTACGGCATT TTGGAATACG ACATCGCACC GCAAACCCGC
30
                    751 GTCCACGCAG GCATGGACTA CCAGCAGGCG AAAGAAACCG CCGACGCGCC
                    801 GCTCAGCTAC GCCGTGTACG ACAGCCAAGG TTATGCCACC GCCTTCGGCC
                          CGAAAGACAA CCCCGCCACA AATTGGGCGA ACAGCCGCCA CCGTGCGCTC
                    851
                         AACCTGTTCG CCGGCATCGA ACACCGCTTC AACCAAGACT GGAAACTCAA
                    901
                    951 AGCCGAATAC GACTACACCC GCAGCCGCTT CCGCCAGCCC TACGGCGTAG
 35
                          CAGGCGTGCT TTCCATCGAC CACAACACCG CCGCCACCGA CCTGATTCCC
                  1051 GGTTATTGGC ACGCCGACCC GCGCACCCAC AGCGCCAGCG TGTCATTGAT
1101 CGGCAAATAC CGCCTGTTCG GCCGCGAACA CGATTTAATC GCGGGTATCA
                   1151 ACGGTTACAA ATACGCCAGC AACAAATACG GCGAACGCAG CATCATCCCC
                   1201 AACGCCATTC CCAACGCCTA CGAATTTTCC CGCACGGGTG CCTACCCGCA
 40
                          GCCTGCATCG TTTGCCCAAA CCATCCCGCA ATACGGCACC AGGCGGCAAA TCGGCGGCTA TCTCGCCACC CGTTTCCGCG CCGCCGACAA CCTTTCGCTG
                   1251
                   1301
                   1351 ATTTTGGGCG GACGATACAC CCGTTACCGC ACCGGCAGCT ACGACAGCCG
                   1401 CACACAAGGC ATGACCTATG TGTCCGCCAA CCGTTTCACC CCCTACACAG
                   1451 GCATCGTGTT CGACCTGACC GGCAACCTGT CTCTTTACGG CTCGTACAGC
1501 AGCCTGTTCG TCCCGCAATC GCAAAAAGAC GAACACGGCA GCTACCTGAA
 45
                   1551 ACCCGTAACC GGCAACAATC TGGAAGCCGG CATCAAAGGC GAATGGCTTG
                   1601 AAGGCCGTCT GAACGCATCC GCCGCCGTGT ACCGCGCCCG TAAAAACAAC
                   1651 CTCGCCACCG CAGCAGGACG CGACCCGAGC GGCAACACCT ACTACCGCGC 1701 CGCCAACCAA GCCAAAACCC ACGGCTGGGA AATCGAAGTC GGCGGCCGCA
 50
                          TCACGCCCGA ATGGCAGATA CAGGCAGGTT ACAGCCAAAG CAAAACCCGC
                   1751
                          GACCAAGACG GCAGCCGCCT GAACCCCGAC AGCGTACCCG AACGCAGCTT
                   1851 CAAACTCTTC ACTGCCTACC ACTTTGCCCC CGAAGCCCCC AGCGGCTGGA
1901 CCATCGGCGC AGGCGTGCGC TGGCAGAGCG AAACCCACAC CGACCCTGCC
                   1951 ACGCTCCGCA TCCCCAACCC CGCCGCCAAA GCCCGCGCCG CCGACAACAG
 55
                          CCGCCAAAAA GCCTACGCCG TCGCCGACAT CATGGCGCGT TACCGCTTCA
                          ATCCGCGCGC CGAACTGTCG CTGAACGTGG ACAATCTGTT CAACAAACAC
                   2051
                          TACCGCACCC AGCCCGACCG CCACAGCTAC GGCGCACTGC GGACAGTGAA
                   2101
                          CGCGGCGTTT ACCTATCGGT TTAAATAA
                   2151
        This corresponds to the amino acid sequence <SEQ ID 666; ORF23-1>:
  60
                          MTRFKYSLLF AALLPVYAQA DVSVSDDPKP QESTELPTIT VTADRTASSN DGYTVSGTHT PLGLPMTLRE IPQSVSVITS QQMRDQNIKT LDRALLQATG
                     101 TSRQIYGSDR AGYNYLFARG SRIANYQING IPVADALADT GNANTAAYER
                     151 VEVVRGVAGL LDGTGEPSAT VNLVRKRLTR KPLFEVRAEA GNRKHFGLDA
```

201 DVSGSLNTEG TLRGRLVSTF GRGDSWRRRE RSRDAELYGI LEYDIAPQTR

		251	VHAGMDYQQA	KETADAPLSY	AVYDSQGYAT	AFGPKDNPAT	NWANSRHRAL
		301	NLFAGIEHRF	NODWKLKAEY	DYTRSRFRQP	YGVAGVLSID	HNTAATDLIP
		351	GYWHADPRTH	SASVSLIGKY	RLFGREHDLI	AGINGYKYAS	NKYGERSIIP
_	•	401	NAIPNAYEFS	RTGAYPQPAS	FAQTIPQYGT	RRQIGGYLAT	RFRAADNLSL
5		451	ILGGRYTRYR				
	•	501	SLFVPQSQKD	EHGSYLKPVT	GNNLEAGIKG	<b>EWLEGRLNAS</b>	AAVYRARKNN
		551					QAGYSQSKTR
		601					WQSETHTDPA
	•	651	TLRIPNPAAK	ARAADNSRQK	AYAVADIMAR	YRFNPRAELS	LNVDNLFNKH
10		701	YRTQPDRHSY	GALRTVNAAF	TYRFK*		

Computer analysis of this amino acid sequence gave the following results:

Homology with the ferric-pseudobactin receptor PupB of *Pseudomonas putida* (accession number P38047) ORF23 and PupB protein show 32% aa identity in 205aa overlap:

```
Orf23
                     FARGSRIANYQINGIPVADALADTGNANTAAYERVEVVRGVAGLLDGTGEPSATVNLVRK 65
15
                          I NY+++G+P + L D + + A ++RVE+VRG GL+ G G PSAT+NL+RK
                 215 WSRGFAIQNYEVDGVPTSTRL-DNYSQSMAMFDRVEIVRGATGLISGMGNPSATINLIRK 273
          PupB
                   RLTRKPLFEVRAEAGNRKHFGLDADVSGSLNTEXXLRGRLVSTFXXXXXXXXXXXXXXXX 125
          Orf23 - 66
                             + EAGN
                                      +G
                                            DVSG L
                                                       +RGR V+ +
20
                 274 RPTAEAQASITGEAGNWDRYGTGFDVSGPLTETGNIRGRFVADYKTEKAWIDRYNQQSQL 333
          PupB
          Orf23
                 126 LYGILEYDIAPQTRVHAXMDYQQAKETADAPLSYAVYD--SQGYATAFGPKDNPATNWAN 183
                                        Y + D+PL +
                                                           SGT
                 334 MYGITEFDLSEDTLLTVGFSY--LRSDIDSPLRSGLPTRFSTGERTNLKRSLNAAPDWSY 391
          PupB
25
          Orf23
                 184 SHHRALNLFAGIEHRFNQDWKLKAE 208
                          + F IE +
          PupB
                 392 NDHEQTSFFTSIEQQLGNGWSGKIE 416
```

#### 30 Homology with a predicted ORF from N. meningitidis (strain A)

ORF23 shows 95.7% identity over a 211aa overlap with an ORF (ORF23a) from strain A of N. meningitidis:

		•		•	10	20	30
25	orf23.pep	•				nyqingipvai	DALADTG
35	F22-	OVERGITTEREDEE				111111111	
	orf23a	QMRDQNIKALDRAL 90	LQATGTSRQ 100	110	YLFARGSRIA 120		
		30	100	110	120	130	140
		40	50	60	. 70	80	90
40	orf23.pep	NANTAAYERVEVVR	GVAGLLDGT	GEPSATVNLV	RKRLTRKPLF	EVRAEAGNRKI	HFGLDAD
•					111 111111	1111111111	1111 11
•	orf23a	NANTAAYERVEVVR					
	•	150	160	170	180	190	200
45		100	110	120	130	140	150
	orf23.pep	VSGSLNTEXXLRGR					
							111111
	orf23a	VSGSLNAEGTLRGR					
50		210	220	230	240	250	260
30	•	160	170	180	190	200	010
÷	orf23.pep	ETADAPLSYAVYDS				200 CTEUDENODWI	210
	01110.606				1:11111111	_	IIIIIII
	orf23a	ETADAPLSYAVYDS	QGYATAFGE				KLKAEYD
55	-	270	280	290	300	310	320
•						2	
	orf23.pep	Y					
	orres.pep	i		•			
60	orf23a	YTRSRFRQPYGVAG	VLSIDHNTA	ATDLIPGYWH	ADPRTHSASV	SLIGKYRLFG	REHDLIA
		330	340	350	360	370	380

The complete length ORF23a nucleotide sequence <SEQ ID 667> is:

	1				GCCGCCCTGT		
	51				CCCAAAACCG		
	101				ACCGCACCGC		
5	151 201				CCGCTCGGGC		
3	251				CATCACATCG		
	301				CCCTGTTGCA GCGGGCTACA		
	351				AATCAACGGC		
	401				ACACCGCCGC		
10	451				CTGGACGGCA		
10	501				CCCGACCCGC		
	551				AACATTTCGG		
	601				ACGCTGCGCG		
	651				GCAGCGCGAA		
15	701				ACATCGCACC		
15	751				AAAGAAACCG		
	801				TTATGCCACC		
	851				ACAGCCGCCA		
	901				AACCAAGACT		
20	951				CCGCCAGCCC		
20	1001				CCGCCACCGA		
	1051				AGCGCCAGCG		
	1101				CGATTTAATC		
	1151				GCGAACGCAG		
25	1201	AACGCCATTC	CCAACGCCTA	CGAATTTTCC	CGCACGGGTG	CCTACCCGCA	
	1251	GCCTGCATCG	TTTGCCCAAA	CCATCCCGCA	ATACGGCACC	AGGCGGCAAA	
	1301				CCGCCGACAA		
	1351				ACCGGCAGCT		
	1401	CACACAAGGC	ATGACCTATG	TGTCCGCCAA	CCGTTTCACC	CCCTACACAG	
30	1451				CGCTTTACGG		
	1501	AGCCTGTTCG	TCCCGCAATC	GCAAAAAGAC	GAACACGGCA	GCTACCTGAA	
	1551	ACCCGTAACC	GGCAACAATC	TGGAAGCCGG	CATCAAAGGC	GAATGGCTTG	
	1601				ACCGCGCCCG		
	1651				GGCAACACCT		
35	1701				AATCGAAGTC		
	1751				ACAGCCAAAG		
	1801	GACCAAGACG	GCAGCCGCCT	GAACCCCGAC	AGCGTACCCG	AACGCAGCTT	
	1851				CGAAGCCCCC		
40	1901				AAACCCACAC		
40	1951				GCCCGCGCCG		
	2001				CATGGCGCGT		
	2051	ATCCGCGCGC	CGAACTGTCG	CTGAACGTGG	ACAATCTGTT	CAACAAACAC	
	2101 2151				GGCGCACTGC	GGACAGTGAA	
	2151	CGCGGCGTTT	ACCTATEGGT	TTAAATAA			
45	This encodes a	orotein havin	o amino ació	1 seguence <	SEO ID 668	>•	
	TIME CHOOCOS W	P	B	- boquenoo	PT C ID 000	•	
	1	MTRFKYSLLF	AAT.T.PVYAOA	DVSVSDDPKP	QESTELPTIT	WP2 ATTACKEN	
	51	DGYTVSGTHT	PLGLPMTLRE	TPOSVSVITES	QQMRDQNIKA	TUBBITOST	
	101				IPVADALADT		
	151	VEVVRGVAGI.	LOGTGEPSAT	UNI.URKRPTR	KPLFEVRAEA	GNRKHEGIGA	
50	201	DVSGSLNAEG	TLRGRLVSTF	GRGDSWRORE	RSRDAELYGI	LEYDIAPOTR	
	251				AFGPKDNPAT		
	301	NLFAGIEHRF	NODWKLKAEY	DYTRSRFROP	YGVAGVLSID	HNTAATDLIP	
	351	GYWHADPRTH	SASVSLIGKY	RLFGREHDLI	AGINGYKYAS	NKYGERSIIP	
	401	NAIPNAYEFS	RTGAYPQPAS	FAQTIPQYGT	RRQIGGYLAT	RFRAADNLSL	
55	451	ILGGRYSRYR	TGSYDSRTQG	MTYVSANRFT	PYTGIVFDLT	GNLSLYGSYS	
	501	SLFVPQSQKD	EHGSYLKPVT	GNNLEAGIKG	<b>EWLEGRLNAS</b>	AAVYRARKNN	
	551	LATAAGRDPS	GNTYYRAANQ	AKTHGWEIEV	GGRITPEWQI	QAGYSQSKTR	
	601	DQDGSRLNPD	SVPERSFKLF	TAYHFAPEAP	SGWTIGAGVR	WQSETHTDPA	
	651	TLRIPNPAAK	ARAADNSRQK	AYAVADIMAR	YRFNPRAELS		
60	701	YRTQPDRHSY	GALRTVNAAF	TYRFK*			
	ODED2s and OI	DECC 1 charre	00.20/ :4	litus in 705 au			
	ORF23a and OF	XF23-1 SHOW	99.2% Iden	nty in 725 aa	i overiap:		
			10	20	30	40 ~	
	orf23a.pe	יים מייות מיי		20 ชอดอยขอยกา	30 PKPQESTELPT:	40 5 TTVTADDTAGG	
	Olizja.pe	11111		11111111111111111111111111111111111111			11111111111111111111111111111111111111
65	orf23-1	MTRFK	YSLLFAALLPV	YAOADVSVSDD	PKPQESTELPT:	ITVTADRTASS	・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・
			10	20		40 5	
						•	

5	orf23a.pep orf23-1	70 80 90 100 110 120 PLGLPMTLREIPQSVSVITSQQMRDQNIKALDRALLQATGTSRQIYGSDRAGYNYLFARG
10	orf23a.pep orf23-1	130 140 150 160 170 180 SRIANYQINGIPVADALADTGNANTAAYERVEVVRGVAGLLDGTGEPSATVNLVRKRPTR
15	orf23a.pep orf23-1	190 200 210 220 230 240  KPLFEVRAEAGNRKHFGLGADVSGSLNAEGTLRGRLVSTFGRGDSWRQRERSRDAELYGI
20	orf23a.pep	250 260 270 280 290 300 LEYDIAPQTRVHAGMDYQQAKETADAPLSYAVYDSQGYATAFGPKDNPATNWANSRHRAL
25	orf23a.pep orf23-1	310 320 330 340 350 360 NLFAGIEHRFNQDWKLKAEYDYTRSRFRQPYGVAGVLSIDHNTAATDLIPGYWHADPRTH
35	orf23a.pep orf23-1	370 380 390 400 410 420 SASVSLIGKYRLFGREHDLIAGINGYKYASNKYGERSIIPNAIPNAYEFSRTGAYPQPAS
40	orf23a.pep orf23-1	430 440 450 460 470 480 FAQTIPQYGTRRQIGGYLATRFRAADNLSLILGGRYSRYRTGSYDSRTQGMTYVSANRFT
45	orf23a.pep	490 500 510 520 530 540 PYTGIVFDLTGNLSLYGSYSSLFVPQSQKDEHGSYLKPVTGNNLEAGIKGEWLEGRLNAS
50	orf23a.pep orf23-1	550 560 570 580 590 600  AAVYRARKNNLATAAGRDPSGNTYYRAANQAKTHGWEIEVGGRITPEWQIQAGYSQSKTR
55	orf23a.pep orf23-1	610 620 630 640 650 660 DQDGSRLNPDSVPERSFKLFTAYHFAPEAPSGWTIGAGVRWQSETHTDPATLRIPNPAAK
65	orf23a.pep orf23-1	670 680 690 700 710 720 ARAADNSRQKAYAVADIMARYRFNPRAELSLNVDNLFNKHYRTQPDRHSYGALRTVNAAF
70	orf23a.pep orf23-1	TYRFKX        TYRFKX

WO 99/24578 PCT/IB98/01665

### Homology with a predicted ORF from N.gonorrhoeae

20

25

ORF23 shows 93.4% identity over a 211aa overlap with a predicted ORF (ORF23.ng) from N. gonorrhoeae:

5	orf23.pep	GYNYLFARGSRIANYQINGIPVADALADTGNANTAAYERVEVVRGVAGLLD	51
_	orf23ng	SAVDACRIPGYNYLFARGSRIANYQINGIPVADALADTGNANTAAYERVEVVRGVAGLPD	60
	orf23.pep	GTGEPSATVNLVRKRLTRKPLFEVRAEAGNRKHFGLDADVSGSLNTEXXLRGRLVSTFGR	111
10	orf23ng	GTGEPSATVNLVRKHPTRKPLFEVRAEAGNRKHFGLGADVSGSLNAEGTLRGRLVSTFGR	120
	orf23.pep	${\tt GDSWRRRERSRXAELYGILEYDIAPQTRVHAXMDYQQAKETADAPLSYAVYDSQGYATAF}$	171
15	orf23ng	:	180
15	orf23.pep	GPKDNPATNWANSHHRALNLFAGIEHRFNQDWKLKAEYDY	211
	orf23ng	GPKDNPATNWSNSRNRALNLFAGIEHRFNQDWKLKAEYDYTRSRFRQPYGVAGVLSIDHS	240

The ORF23ng nucleotide sequence <SEQ ID 669> is predicted to encode a protein comprising amino acid sequence <SEQ ID 670>:

```
1 SAVDACRIPG YNYLFARGSR IANYQINGIP VADALADTGN ANTAAYERVE
                      VVRGVAGLPD GTGEPSATVN LVRKHPTRKP LFEVRAEAGN RKHFGLGADV
                 101 SGSLNAEGTL RGRLVSTFGR GDSWRQLERS RDAELYGILE YDIAPQTRVH
                 151 AGMDYQQAKE TADAPLSYAV YDSQGYATAF GPKDNPATNW SNSRNRALNL
25
                 201 FAGIEHRFNQ DWKLKAEYDY TRSRFRQPYG VAGVLSIDHS TAATDLIPGY
                      WHADPRTHSA SMSLTGKYRL FGREHDLIAG INGYKYASNK YGERSIIPNA IPNAYEFSRT GAYPQPSSFA QTIPQYDTRR QIGGYLATRF RAADNLSLIL
                 251
                 301
                 351 GGRYSRYRAG SYNSRTQGMT YVSANRFTPY TGIVFDLTGN LSLYGSYSSL
                      FVPQLQKDEH GSYLKPVTGN NLEADIKGEW LEGRLNASAA VYRARKNNLA
                 401
                 451 TAAGRDQSGN TYYRAANQAK THGWEIEVGG RITPEWQIQA GYSQSKPRDQ
30
                 501 DGSRLNPDSV PERSFKLFTA YHLAPEAPSG RTIGAGVRRQ GETHTDPAAL
                 551 RIPNPAAKAR AVANSRQKAY AVADIMARYR FNPRTELSLN VDNLFNKHYR
                 601 TOPDRHSYGA LRTVNAAFTY RFK*
```

## Further work revealed the complete nucleotide sequence <SEQ ID 671>:

35	1	ATGACACGCT	TCAAATACTC	CCTGCTTTTT	GCCGCCCTGC	TACCCGTGTA
	51	CGCGCAGGCC	GATGTTTCTG	TTTCAGACGA	CCCCAAACCG	CAGGAAAGCA
	101	CCGAATTGCC	GACCATCACC	GTTACCGCCG	ACCGCACCGC	GAGTTCCAAC
	151	GACGGCTACA	CCGTTTCCGG	CACGCACACC	CCGTTCGGGC	TGCCCATGAC
	201	CCTGCGCGAA	ATCCCGCAGA	GCGTCAGCGT	CATCACATCG	CAACAAATGC
40	251	GCGACCAAAA	CATCAAAACG	CTCGACCGCG	CCCTGTTGCA	GGCGACCGGC
	301	ACCAGCCGCC	AGATTTACGG	CTCCGACCGC	GCGGGCTACA	ACTACCTGTT
	351	CGCGCGCGGC	AGCCGCATCG	CCAACTACCA	AATCAACGGC	ATCCCCGTTG
	401	CCGACGCGCT	GGCCGATACG	GGCAATGCCA	ACACCGCCGC	CTATGAGCGC
	451			GGCGGGGCTG		
45	501	TTCTGCCACC	GTCAATCTGG	TACGCAAACA	CCCGACCCGC	AAGCCATTGT
	551	TTGAAGTCCG	CGCCGAAGCC	GGCAACCGCA	AACATTTCGG	GCTGGGCGCG
	601	GACGTATCGG	GCAGCCTGAA	CGCCGAAGGC	ACGCTGCGCG	GCCGCCTGGT
	651	TTCCACCTTC	GGACGCGGCG	ACTCGTGGCG	GCAGCTCGAA	CGCAGCCGCG
	701	ATGCCGAACT	CTACGGCATT	TTGGAATACG	ACATCGCACC	<b>GCAAACCCGC</b>
50	751	GTCCACGCAG	GCATGGACTA	CCAGCAGGCG	AAAGAAACCG	CAGACGCGCC
	801	GCTCAGCTAC	GCCGTGTACG	ACAGCCAAGG	TTATGCCACC	GCCTTCGGCC
	851	CAAAAGACAA	CCCCGCCACA	<b>AATTGGTCGA</b>	ACAGCCGCAA	CCGTGCGCTC
	901	AACCTGTTCG	CCGGCATAGA	ACACCGCTTC	AACCAAGACT	<b>GGAAACTCAA</b>
	951	AGCCGAATAC	GACTACACCC	GTAGCCGCTT	CCGCCAGCCC	TACGGTGTGG
55	1001	CAGGCGTACT	TTCCATCGAC	CACAGCACTG	CCGCCACCGA	CCTGATTCCC
	1051	GGTTATTGGC	ACGCcgatcc	GCGCACCCAC	AGCGCCAGCA	TGTCATTGAC
	1101	CGGCAAATAC	CgcctGTTCG	GCCGCGAGCA	CGATTTAATC	GCGGGTATCA
	1151	ACGGCTACAA	ATACGCCAGC	AACAAATACG	GCGAACGCAG	CATCATTCCC
	1201	AACGCCATTC	CCAACGCCTA	CGAATTTTCC	CGCACGGGCG	CCTATCCGCA
60	1251	GCCATCATCG	TTTGCCCAAA	CCATCCCGCA	ATACGACACC	AGGCGGCAAA
	1301	TCGGCGGCTA	TCTCGCCACC	CGTTTCCGCG	CCGCCGACAA	CCTTTCGCTG
	1351	ATACTCGGCG	GCAGATACAG	CCGCTACCGC	GCAGGCAGCT	ACAACAGCCG

	1401 CACA	CAAGGC .	ATGACCTATG	TGTCCGCCAA	CCGTTTCACC	CCCTACACAG	
	1451 GCAT	CGTGTT	CGATCTGACC	GGCAACCTGT	CGCTTTACGG	CTCGTACAGC	
					GAACACGGCA		
5					CATCAAAGGC		
3	1601 AAGG	GCGTCT	GAACGCATCC	GCCGCCGTGT	ACCGCGCCCG	TAAAAACAAC	
	1651 CTCG 1701 CGCC	DOCAD	CAGCAGGACG	ACCCOMCCCA	GGCAACACCT AATCGAAGTC	ACTATCGCGC	
	1751 CGCC	CCCCCA	ATCCCACATA	CACCCACCCT	ACAGCCAAAG	CARACCCCCC	•
	1801 GACC	ADCACE	CCACCCCCCT	CAGGCAGGCI	AGCGTACCCG	AACCCACCE	
10	1851 CAAA	CTCTTC	ACCGCCTACC	ACTTACCCCAC	CGAAGCCCCC	AGCGGCCGGA	
					AAACCCACAC		
					GCCCGCGCCG		
					CATGGCGCGT		
	2051 ATCC	GCGCAC	CGAACTGTCG	CTGAACGTGG	ACAACCTGTT	CAACAAACAC	
15	2101 TACC	GCACCC	AGCCCGACCG	CCACAGCTAC	GGCGCACTGC	GGACAGTGAA	
			ACCTATCGGT		•	•	
	(D) 1	L•	• <b>1</b>	-070 H	0 (50 ODE)		
	This corresponds to the	ne amino	o acid seque	nce <seq 11<="" th=""><th>D 672; ORF2</th><th>23ng-1&gt;:</th><th></th></seq>	D 672; ORF2	23ng-1>:	
					QESTELPTIT		
20	101 TSRQ	TVCCDD	A COUNT FARC	TPQSVSVITS	QQMRDQNIKT	LDRALLQATG	
20	151 VEVV	DCAPCI	DUCACEDENA	SKIANIQING	I PVADALADT KPLFEVRAEA	GNANTAAYER	
	201 DVSG	SINAEC	TIDCHESKI	CDCDCMDOIL	RSRDAELYGI	TEVETABORD	•
	251 VHAG	ACCYCL	KETADAPI.SY	AVYDSOCYAT	AFGPKDNPAT	PEIDIWEAT	
	301 NLFA	GIEHRF	NODWKLKAEY	DYTESEFROP	YGVAGVLSID	HSTADMINAL	· .
25	351 GYWH	ADPRTH	SASMSLTGKY	RLFGREHDLI	AGINGYKYAS	NKYGERSIIP	,
	401 NAIP	NAYEFS	RTGAYPQPSS	FAQTIPQYDT	RRQIGGYLAT	RFRAADNLSL	,
					PYTGIVFDLT		
	501 SLFV	PQLQKD	EHGSYLKPVT	GNNLEADIKG	<b>EWLEGRLNAS</b>	AAVYRARKNN	
••	551 LATA	AGRDQS	GNTYYRAANQ	AKTHGWEIEV	GGRITPEWQI	QAGYSQSKPR	
30	601 DQDG	SRLNPD	SVPERSFKLF	TAYHLAPEAP	SGRTIGAGVR	RQGETHTDPA	•
-					YRFNPRTELS	LNVDNLFNKH	
	701 YRTQ	PDRHSY	GALRTVNAAF	TYRFK*			
	ORF23ng-1 and ORF	23-1 sh	ow 95.9% id	lentity in 72:	5 aa overlap:		
				-	A .		
	•				•		
25		\mp.=	10	20	30	40 5	0 60
35	orf23-1.pep	MTRFKY	SLLFAALLPV	YAQADVSVSDD	30 PKPQESTELPT	ITVTADRTASS	NDGYTVSGTHT
35		111111	SLLFAALLPVY	YAQADVSVSDDI 	30 PKPQESTELPT	ITVTADRTASS	NDGYTVSGTHT
35	orf23-1.pep orf23ng-1	111111	SLLFAALLPVY             SLLFAALLPVY	(AQADVSVSDD)             (AQADVSVSDD)	30 PKPQESTELPT          PKPQESTELPT	ITVTADRTASS              ITVTADRTASS	NDGYTVSGTHT
35		111111	SLLFAALLPVY	YAQADVSVSDDI 	30 PKPQESTELPT          PKPQESTELPT	ITVTADRTASS            ITVTADRTASS	NDGYTVSGTHT
35		111111	SLLFAALLPVY           SLLFAALLPVY 10	(AQADVSVSDD)              AQADVSVSDD)   20	30 PKPQESTELPT           PKPQESTELPT	ITVTADRTASS              TVTADRTASS 40 5	NDGYTVSGTHT              NDGYTVSGTHT 0 60
		 MTRFKY	SLLFAALLPVY 	YAQADVSVSDD:             YAQADVSVSDD: 20	30 PKPQESTELPT           PKPQESTELPT 30  90 1	ITVTADRTASS            ITVTADRTASS 40 5	NDGYTVSGTHT            NDGYTVSGTHT 0 60
	orf23ng-1 orf23-1.pep	MTRFKY PLGLPM	SLLFAALLPVY	YAQADVSVSDD:             YAQADVSVSDD: 20   80 VITSQQMRDQN:	30 PKPQESTELPT           PKPQESTELPT 30 90 10 IKTLDRALLQA	ITVTADRTASS            ITVTADRTASS 40 5 00 11 TGTSRQIYGSD	NDGYTVSGTHT           NDGYTVSGTHT 0 60 0 120 PRAGYNYLFARG
	orf23ng-1	PLGLPM	SLLFAALLPV)	YAQADVSVSDD:             YAQADVSVSDD: 20   80    TSQQMRDQN: 	30 PKPQESTELPT           PKPQESTELPT 30 90 10 IKTLDRALLQA	ITVTADRTASS	NDGYTVSGTHT           NDGYTVSGTHT 0 60 0 120 PRAGYNYLFARG
40	orf23ng-1 orf23-1.pep	PLGLPM	SLLFAALLPV)	YAQADVSVSDD:             YAQADVSVSDD: 20   80    TSQQMRDQN: 	30 PKPQESTELPT           PKPQESTELPT 30  90 10 IKTLDRALLQA	ITVTADRTASS	NDGYTVSGTHT           NDGYTVSGTHT 0 60 0 120 RAGYNYLFARG
	orf23ng-1 orf23-1.pep	PLGLPM	SLLFAALLPVY IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	YAQADVSVSDD:            YAQADVSVSDD: 20  80 VITSQQMRDQN:            VITSQQMRDQN:	30 PKPQESTELPT           PKPQESTELPT 30  90 10 IKTLDRALLQA	ITVTADRTASS            ITVTADRTASS 40 5 00 11 TGTSRQIYGSD	NDGYTVSGTHT           NDGYTVSGTHT 0 60 0 120 RAGYNYLFARG
40	orf23ng-1 orf23-1.pep orf23ng-1	PLGLPM	SLLFAALLPVY IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	YAQADVSVSDD:            YAQADVSVSDD: 20  80  YITSQQMRDQN:           YITSQQMRDQN: 80  140	30 PKPQESTELPT           PKPQESTELPT 30  90 1 IKTLDRALLQA            IKTLDRALLQA 90 1	ITVTADRTASS	NDGYTVSGTHT           NDGYTVSGTHT 0 60  0 120 RAGYNYLFARG            RAGYNYLFARG 0 120  0 180
40	orf23ng-1 orf23-1.pep	PLGLPM 1:     PFGLPM SRIANY	SLLFAALLPVY 111111111111111111111111111111111111	YAQADVSVSDD:            YAQADVSVSDD: 20  80  YITSQQMRDQN:            YITSQQMRDQN: 80  140  LADTGNANTAA	30 PKPQESTELPT           PKPQESTELPT 30  90 1 IKTLDRALLQA              IKTLDRALLQA 90 1 IKTLDRALLQA	ITVTADRTASS	NDGYTVSGTHT           NDGYTVSGTHT 0 60  0 120 RAGYNYLFARG            RAGYNYLFARG 0 120  0 180 TVNLVRKRLTR
40	orf23ng-1 orf23-1.pep orf23ng-1 orf23-1.pep	PLGLPM 1:    PFGLPM SRIANY	SLLFAALLPVY 111111111111111111111111111111111111	YAQADVSVSDD:            YAQADVSVSDD: 20  80  VITSQQMRDQN:            VITSQQMRDQN: 80  140  LADTGNANTAA	30 PKPQESTELPT           PKPQESTELPT 30  90 1 IKTLDRALLQA            IKTLDRALLQA 90 1 150 1 YERVEVVRGVA	ITVTADRTASS	NDGYTVSGTHT            NDGYTVSGTHT 0 60  0 120 RAGYNYLFARG           RAGYNYLFARG 0 120  0 180 TVNLVRKRLTR
40 45	orf23ng-1 orf23-1.pep orf23ng-1	PLGLPM 1:    PFGLPM SRIANY	SLLFAALLPVY 111111111111111111111111111111111111	YAQADVSVSDD:            YAQADVSVSDD: 20  80  YITSQQMRDQN:            YITSQQMRDQN: 80  140  LADTGNANTAA:	30 PKPQESTELPT           PKPQESTELPT 30  90 1 IKTLDRALLQA            IKTLDRALLQA 10 150 1 YERVEVVRGVA	ITVTADRTASS	NDGYTVSGTHT            NDGYTVSGTHT 0 60  0 120 RAGYNYLFARG            RAGYNYLFARG 0 120  0 180 TVNLVRKRLTR       :
40	orf23ng-1 orf23-1.pep orf23ng-1 orf23-1.pep	PLGLPM 1:    PFGLPM SRIANY	SLLFAALLPVY 111111111111111111111111111111111111	YAQADVSVSDD:            YAQADVSVSDD: 20  80  YITSQQMRDQN:            YITSQQMRDQN: 80  140  LADTGNANTAA:	30 PKPQESTELPT           PKPQESTELPT 30  90 1 IKTLDRALLQA            IKTLDRALLQA 10 150 1 YERVEVVRGVA	ITVTADRTASS	NDGYTVSGTHT            NDGYTVSGTHT 0 60  0 120 RAGYNYLFARG            RAGYNYLFARG 0 120  0 180 TVNLVRKRLTR       :
40 45	orf23ng-1 orf23-1.pep orf23ng-1 orf23-1.pep	PLGLPM 1:    PFGLPM SRIANY	SLLFAALLPVY 111111111111111111111111111111111111	YAQADVSVSDD:            YAQADVSVSDD: 20  80  VITSQQMRDQN:           VITSQQMRDQN: 80  140  LADTGNANTAA:            LADTGNANTAA:	30 PKPQESTELPT           PKPQESTELPT 30  90 1 IKTLDRALLQA            IKTLDRALLQA 90 1  150 1 YERVEVVRGVA            YERVEVVRGVA	ITVTADRTASS	NDGYTVSGTHT
40 45	orf23ng-1 orf23-1.pep orf23ng-1 orf23-1.pep orf23ng-1	PLGLPM  :     PFGLPM  SRIANY	SLLFAALLPVY 111111111111111111111111111111111111	AQADVSVSDD:             AQADVSVSDD:             AQADVSVSDD:               VITSQQMRDQN:               VITSQQMRDQN:               ADTGNANTAA:              LADTGNANTAA:              200	30 PKPQESTELPT           PKPQESTELPT 30  90 1 IKTLDRALLQA            IKTLDRALLQA            IKTLDRALLQA 150 1 YERVEVVRGVA            YERVEVVRGVA            YERVEVVRGVA	ITVTADRTASS	NDGYTVSGTHT
40 45 50	orf23ng-1 orf23-1.pep orf23ng-1 orf23-1.pep	PLGLPM  :     PFGLPM  :     SRIANY  :     SRIANY	SLLFAALLPV                          SLLFAALLPV  10  70  ITLREIPQSVSV  70  130  IQINGIPVADAI                     QUINGIPVADAI 130  190  TABEAGNRKHFO	AQADVSVSDD:             AQADVSVSDD:             AQADVSVSDD:               VITSQQMRDQN:              VITSQQMRDQN:              VITSQQMRDQN:              ADTGNANTAA:              LADTGNANTAA:              200  SLDADVSGSLN:	30 PKPQESTELPT           PKPQESTELPT 30  90 1 IKTLDRALLQA            IKTLDRALLQA            IKTLDRALLQA 150 1 YERVEVVRGVA            YERVEVVRGVA 150 1	ITVTADRTASS	NDGYTVSGTHT           NDGYTVSGTHT 0 60  0 120 RAGYNYLFARG            RAGYNYLFARG 0 120  0 180 TVNLVRKRLTR            TVNLVRKHPTR 0 180  0 240 ERSRDAELYGI
40 45	orf23ng-1 orf23-1.pep orf23ng-1 orf23-1.pep orf23ng-1	PLGLPM  :     PFGLPM  :     SRIANY  :     SRIANY	SLLFAALLPV                          SLLFAALLPV  10  70  TITREIPQSVSV  70  130  CQINGIPVADAI                     CQINGIPVADAI 130  190  TRAEAGNRKHFO	AQADVSVSDD:             AQADVSVSDD:             AQADVSVSDD:	30 PKPQESTELPT           PKPQESTELPT 30  90 10 IKTLDRALLQA             IKTLDRALLQA 90 10 150 10 YERVEVVRGVA             YERVEVVRGVA             YERVEVVRGVA             YERVEVVRGVA             YERVEVVRGVA	ITVTADRTASS             ITVTADRTASS 40 5  00 11 TGTSRQIYGSD            TGTSRQIYGSD 00 11  60 17 GLLDGTGEPSA            GLPDGTGEPSA 60 17  20 23 TFGRGDSWRRR	NDGYTVSGTHT           NDGYTVSGTHT 0 60  0 120 RAGYNYLFARG            RAGYNYLFARG 0 120  0 180 TVNLVRKRLTR           TVNLVRKHPTR 0 180  0 240 ERSRDAELYGI
40 45 50	orf23ng-1  orf23-1.pep orf23-1.pep orf23ng-1  orf23-1.pep	PLGLPM  :     PFGLPM  :     SRIANY  :     SRIANY	SLLFAALLPV                          SLLFAALLPV  10  70  TITREIPQSVSV  70  130  CQINGIPVADAI                     CQINGIPVADAI 130  190  TRAEAGNRKHFO	AQADVSVSDD:             AQADVSVSDD:             AQADVSVSDD:               VITSQQMRDQN:               VITSQQMRDQN:                ADTGNANTAA:              LADTGNANTAA:              200  SLDADVSGSLN:             GLGADVSGSLN:	30 PKPQESTELPT           PKPQESTELPT 30  90 10 IKTLDRALLQA            IKTLDRALLQA             IKTLDRALLQA 150 1 YERVEVVRGVA            YERVEVVRGVA 150 1 210 2 TEGTLRGRLVS :          AEGTLRGRLVS	ITVTADRTASS	NDGYTVSGTHT           NDGYTVSGTHT 0 60  0 120 RAGYNYLFARG            RAGYNYLFARG 0 120  0 180 TVNLVRKRLTR            TVNLVRKHPTR 0 180  0 240 ERSRDAELYGI
40 45 50	orf23ng-1  orf23-1.pep orf23-1.pep orf23ng-1  orf23-1.pep	PLGLPM  :     PFGLPM  :     SRIANY  :     SRIANY	SLLFAALLPV                          SLLFAALLPV  10  70  TTLREIPQSVSV  70  130  QUINGIPVADAI                     QUINGIPVADAI 130  190  TRAEAGNRKHFO	AQADVSVSDD:             AQADVSVSDD:             AQADVSVSDD:               VITSQQMRDQN:               VITSQQMRDQN:                ADTGNANTAA:              LADTGNANTAA:              200  SLDADVSGSLN:             GLGADVSGSLN:	30 PKPQESTELPT           PKPQESTELPT 30  90 1 IKTLDRALLQA            IKTLDRALLQA             IKTLDRALLQA 150 1 YERVEVVRGVA            YERVEVVRGVA            YERVEVVRGVA            STEGTLRGRLVS            AEGTLRGRLVS	ITVTADRTASS	NDGYTVSGTHT           NDGYTVSGTHT 0 60  0 120 RAGYNYLFARG            RAGYNYLFARG 0 120  0 180 TVNLVRKRLTR            TVNLVRKHPTR 0 180  0 240 ERSRDAELYGI
40 45 50	orf23ng-1  orf23-1.pep orf23ng-1  orf23-1.pep orf23ng-1  orf23-1.pep orf23ng-1	PLGLPM  :     PFGLPM  SRIANY        SRIANY  KPLFEV	SLLFAALLPVY  IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	AQADVSVSDD:             AQADVSVSDD:             AQADVSVSDD:                VITSQQMRDQN:                VITSQQMRDQN:               ADTGNANTAA:               LADTGNANTAA:               LADTGNANTAA:               LADTGNANTAA:               LADTGNANTAA:               LADTGNANTAA:                LADTGNANTAA:                 LADTGNANTAA:                LADTGNANTAA:                 LADTGNANTAA:                  LADTGNANTAA:                   LADTGNANTAA:	30 PKPQESTELPT           PKPQESTELPT 30  90 10 IKTLDRALLQA             IKTLDRALLQA             IKTLDRALLQA              PKPVEVVRGVA              YERVEVVRGVA             YERVEVVRGVA             YERVEVVRGVA             YERVEVVRGVA             YERVEVVRGVA              YERVEVVRGVA              YERVEVVRGVA              YERVEVVRGVA               YERVEVVRGVA                YERVEVVRGVA                YERVEVVRGVA                 YERVEVVRGVA                YERVEVVRGVA                 YERVEVVRGVA	ITVTADRTASS	NDGYTVSGTHT
40 45 50 55	orf23ng-1  orf23-1.pep orf23-1.pep orf23ng-1  orf23-1.pep	PLGLPM I:       PFGLPM SRIANY IIIII SRIANY KPLFEV          KPLFEV	SLLFAALLPVY  IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	AQADVSVSDD:             AQADVSVSDD:             AQADVSVSDD:               VITSQQMRDQN:               VITSQQMRDQN:               ADTGNANTAA:               LADTGNANTAA:               LADTGNANTAA:               LADTGNANTAA:               LADTGNANTAA:               LADTGNANTAA:                LADTGNANTAA:                LADTGNANTAA:                LADTGNANTAA:                LADTGNANTAA:                 LADTGNANTAA:                 LADTGNANTAA:                   LADTGNANTAA:	30 PKPQESTELPT           PKPQESTELPT 30  90 10 IKTLDRALLQA             IKTLDRALLQA             IKTLDRALLQA             IKTLDRALLQA             YERVEVVRGVA             YERVEVVRGVA              YERVEVVRGVA             YERVEVVRGVA             YERVEVVRGVA             YERVEVVRGVA              YERVEVVRGVA             YERVEVVRGVA              YERVEVVRGVA               YERVEVVRGVA                YERVEVVRGVA                YERVEVVRGVA               YERVEVVRGVA                YERVEVVRGVA                 YERVEVVRGVA	ITVTADRTASS	NDGYTVSGTHT
40 45 50	orf23ng-1  orf23-1.pep orf23ng-1  orf23-1.pep orf23ng-1  orf23-1.pep orf23ng-1	PLGLPM  :     PFGLPM  :     SRIANY        KPLFEV        KPLFEV	SLLFAALLPVY  IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	AQADVSVSDD:             AQADVSVSDD:             AQADVSVSDD:               BO    ITSQQMRDQN:               ADTGNANTAA:               ADTGNANTAA:                ADTGNANTAA:                ADTGNANTAA:                ADTGNANTAA:                ADTGNANTAA:                ADTGNANTAA:                 ADTGNANTAA:                 ADTGNANTAA:                  ADTGNANTAA:                  ADTGNANTAA:	30 PKPQESTELPT           PKPQESTELPT 30  90 10 IKTLDRALLQA            IKTLDRALLQA 90 10  150 10 YERVEVVRGVA            YERVEVVRGVA 150 10 210 20 TEGTLRGRLVS            AEGTLRGRLVS 210 20 270 20 LSYAVYDSQGY	ITVTADRTASS	NDGYTVSGTHT
40 45 50 55	orf23ng-1  orf23-1.pep orf23ng-1  orf23-1.pep orf23ng-1  orf23-1.pep orf23ng-1	PLGLPM  :     PFGLPM  :     SRIANY        KPLFEV        KPLFEV	SLLFAALLPVY  IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	AQADVSVSDD:             AQADVSVSDD:             AQADVSVSDD:               AQADVSVSDD:                VITSQQMRDQN:                ADTGNANTAA:                ADTGNANTAA:                ADTGNANTAA:                ADTGNANTAA:                ADTGNANTAA:                ADTGNANTAA:                ADTGNANTAA:                  ADTGNANTAA:                 ADTGNANTAA:                 AUTGNANTAA:                  AUTGNANTAA:                  AUTGNANTAA:                   AUTGNANTAA:                   AUTGNANTAA:	30 PKPQESTELPT            PKPQESTELPT 30  90 10 IKTLDRALLQA            IKTLDRALLQA 90 10  150 10 YERVEVVRGVA 150 10 YERVEVVRGVA 150 10 ZERTLRGRLVS            AEGTLRGRLVS 210 20 ZERTLRGRLVS 210 20 ZERTLRG	ITVTADRTASS	NDGYTVSGTHT
40 45 50	orf23ng-1  orf23-1.pep orf23ng-1  orf23-1.pep orf23ng-1  orf23-1.pep orf23ng-1	PLGLPM  :     PFGLPM  :     SRIANY        KPLFEV        KPLFEV	SLLFAALLPVY  IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	AQADVSVSDD:             AQADVSVSDD:             AQADVSVSDD:               VITSQQMRDQN:                VITSQQMRDQN:                ADTGNANTAA:                ADTGNANTAA:                ADTGNANTAA:                ADTGNANTAA:                ADTGNANTAA:                ADTGNANTAA:                 ADTGNANTAA:                 ADTGNANTAA:                 ADTGNANTAA:                  ADTGNANTAA:                  AUTGNANTAA:                  AUTGNANTAA:                   AUTGNANTAA:                   AUTGNANTAA:	30 PKPQESTELPT            PKPQESTELPT 30  90 10 IKTLDRALLQA            IKTLDRALLQA 90 10  150 10 YERVEVVRGVA            YERVEVVRGVA             YERTUEVVRGVA             YERTUEVVRGVA             YERTUEVVRGVA             YERTUEVVRGVA             YERTUEVVRGVA             YERTUEVVRGVA              YERTUEVVRGVA              YERTUEVVRGVA               YERTUEVVRGVA                YERTUEVVRGVA                YERTUEVVRGVA                YERTUEVVRGVA	ITVTADRTASS	NDGYTVSGTHT
40 45 50 55	orf23ng-1  orf23-1.pep orf23ng-1  orf23-1.pep orf23ng-1  orf23-1.pep orf23ng-1	PLGLPM  :     PFGLPM  :     SRIANY        KPLFEV        KPLFEV	SLLFAALLPVY  IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	AQADVSVSDD:              AQADVSVSDD:              AQADVSVSDD:               VITSQQMRDQN:               VITSQQMRDQN:               VITSQQMRDQN:               ADTGNANTAA:              LADTGNANTAA:              LADTGNANTAA:              LADTGNANTAA:              LADTGNANTAA:              LADTGNANTAA:              LADTGNANTAA:              LADTGNANTAA:              LADTGNANTAA:              LADTGNANTAA:               LADTGNANTAA:               LADTGNANTAA:               LADTGNANTAA:                LADTGNANTAA:                 LADTGNANTAA:                 LADTGNANTAA:                LADTGNANTAA:                 LADTGNANTAA:                 LADTGNANTAA:                 LADTGNANTAA:                 LADTGNANTAA:                  LADTGNANTAA:	30 PKPQESTELPT           PKPQESTELPT 30  90 10 IKTLDRALLQA            IKTLDRALLQA             IKTLDRALLQA 10 150 10 YERVEVVRGVA             YERVEVVRGVA 150 10 210 20 TEGTLRGRLVS             AEGTLRGRLVS 210 20 LSYAVYDSQGY             LSYAVYDSQGY 270 20	ITVTADRTASS	NDGYTVSGTHT
40 45 50 55	orf23ng-1  orf23-1.pep orf23ng-1  orf23-1.pep orf23ng-1  orf23-1.pep orf23ng-1  orf23-1.pep	PLGLPM  :     PFGLPM  :     SRIANY        SRIANY        KPLFEV        LEYDIA	SLLFAALLPV	AQADVSVSDD:             AQADVSVSDD:              AQADVSVSDD:               ADADVSVSDD:               ITSQQMRDQN:               ITSQQMRDQN:               ADTGNANTAA:              LADTGNANTAA:               LADTGNANTAA:               LADTGNANTAA:               LADTGNANTAA:               LADTGNANTAA:                  LADTGNANTAA:                 LADTGNANTAA:	30 PKPQESTELPT:              PKPQESTELPT: 30  90 10 IKTLDRALLQA:             IKTLDRALLQA: 90 10  150 10 YERVEVVRGVA:            YERVEVVRGVA:             AEGTLRGRLVS:             AEGTLRGRLVS:             AEGTLRGRLVS:             AEGTLRGRLVS:             LSYAVYDSQGY:              LSYAVYDSQGY:              LSYAVYDSQGY:              LSYAVYDSQGY:               LSYAVYDSQGY:	ITVTADRTASS	NDGYTVSGTHT
40 45 50 55	orf23ng-1  orf23-1.pep orf23ng-1  orf23-1.pep orf23ng-1  orf23-1.pep orf23ng-1	PLGLPM  :     PFGLPM  :     SRIANY        SRIANY        KPLFEV        LEYDIA	SLLFAALLPVY  IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	AQADVSVSDD:             AQADVSVSDD:              AQADVSVSDD:               ADATGMANTAA:               ADTGMANTAA:               ADTGMANTAA:               CADDVSGSLM:              CADDVSGSLM:              CADDVSGSLM:              CADDVSGSLM:              CADDVSGSLM:              CADDVSGSLM:              CADDVSGSLM:             CADDVSGSLM:              CADDVSGSLM:              CADDVSGSLM:              CADDVSGSLM:              CADDVSGSLM:              CADDVSGSLM:              CADDVSGSLM:              CADDVSGSLM:              CADDVSGSLM:              CADDVSGSLM:              CADDVSGSLM:               CADDVSGSLM:               CADDVSGSLM:               CADDVSGSLM:                CADDVSGSLM:                CADDVSGSLM:                 CADDVSGSLM:                   CADDVSGSLM:                    CADDVSGSLM:	30 PKPQESTELPT           PKPQESTELPT 30  90 10 IKTLDRALLQA            IKTLDRALLQA 90 10  150 10 YERVEVVRGVA            YERVEVVRGVA             YERVEVVRGVA            YERVEVVRGVA            YERVEVVRGVA            YERVEVVRGVA            YERVEVVRGVA            YERVEVVRGVA             YERVEVVRGVA             YERVEVVRGVA             YERVEVVRGVA             YERVEVVRGVA             YERVEVVRGVA	ITVTADRTASS	NDGYTVSGTHT
40 45 50 55	orf23ng-1  orf23-1.pep orf23ng-1  orf23-1.pep orf23ng-1  orf23-1.pep orf23ng-1  orf23-1.pep	PLGLPM  :     PFGLPM  :     SRIANY        SRIANY        KPLFEV        LEYDIA	SLLFAALLPVY  IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	AQADVSVSDD:             AQADVSVSDD:              AQADVSVSDD:               ADATGNANTAA:                ADTGNANTAA:                ADTGNANTAA:                COO    COO	30 PKPQESTELPT           PKPQESTELPT 30  90 10 IKTLDRALLQA            IKTLDRALLQA             IKTLDRALLQA 150 10 YERVEVVRGVA             YERVEVVRGVA             YERVEVVRGVA             YERVEVVRGVA             YERVEVVRGVA             YERVEVVRGVA             YERVEVVRGVA              YERVEVVRGVA	ITVTADRTASS	NDGYTVSGTHT

310

WO 99/24578 PCT/IB98/01665

320 330 340

350

360

		310	320	330	340	350	360
		370	380	390	400	410	420
5	orf23-1.pep	SASVSLIGKYRLFG					
3	orf23ng-1	:          SASMSLTGKYRLFG					
	01200mg 0	370	380	390	400	410	420
		430	440	450	460	470	480
10	orf23-1.pep	FAQTIPQYGTRRQI					
	orf23ng-1		CCVI AMDERIA	111111111	:   :	:	
	0f123lig=1	430	440	450	460	470	480
15		490	E00	. 610	500	500	5.40
13	orf23-1.pep	PYTGIVFDLTGNLS	500 LYGSYSSLFV	510 POSOKDEHGS	520 YLKPVTGNNL	530 EAGIKGEWL	540 EGRLNAS
	500		1111111111	31 113111	1111111111	11 111111	111111
	orf23ng-1	PYTGIVFDLTGNLS 490	500	PQLQKDEHGS 510	YLKPVTGNNL 520	EADIKGEWLI 530	EGRLNAS 540
20							
	orf23-1.pep	550 AAVYRARKNNLATA	560 AGRDPSGNTY	570 YYRAANOAKTH	580 IGWETEVGGRT	590	600 YSOSKTB
			1111 11111		11111111111	11111111	111111
25	orf23ng-1	AAVYRARKNNLATA 550	AGRDQSGNTY 560	YRAANQAKTH 570	IGWEIEVGGRI 580	TPEWQIQAG 590	YSQSKPR 600
				_	300	370	000
	orf23-1.pep	610 DQDGSRLNPDSVPE	620 PSEKI.ETDVI	630 IFA DE A DECWT	640 TCACUDWOSE	650 מ זשת החתם ה	660
••		11111111111111111	1111111111	:::::::::::::::::::::::::::::::::::::::	111111 1:1	111111:11	111111
30	orf23ng-1	DQDGSRLNPDSVPE 610	RSFKLFTAYI 620	ILAPEAPSGRT 630	'IGAGVRRQGE 640	THTDPAALR	IPNPAAK 660
		010		030	040	650	660
	orf23-1.pep	670 ARAADNSRQKAYAV	680	690	700	710	720
35	01123-1.pep			:			KIVNAAF 
	orf23ng-1	ARAVANSRQKAYAV 670				PDRHSYGAL	RTVNAAF
		070	000	090	700	710	720
40	orf23-1.pep	TYRFKX					
40							
	orf23ng-1	TYRFKX					
	In addition, ORF23ng	g-1 shows signification	cant homol	ogy with an	OMP from	E.coli:	
45	sp P16869 FHUE FERRIOXAMINE B	_ECOLI OUTER-N AND FE(III)-RHO	MEMBRANE I	RECEPTOR I	FOR FE(III RSOR >ail16	[)-COPROGE 51542 (an)	N, FE(III)- PIDId1015403
	(D90745) Out	er membrane	protein	FhuE p	recursor	[Escheri	chia colil
	>gi 1651545 gn (Escherichia d	l PID d1015405 coli] >gi 17873	(D90746) 44 (AE000)	Outer me	mbrane pr -membrane	otein Fhu receptor	E precursor
50	coprogen, Fe	(III)-ferrioxam	ine B a	nd Fe(II	I)-rhodotrı	ilic acid	d precursor
50		oli) Length = 7 bits (843), Exp		90			
		228/717 (31%),			(48%), Gar	s = 60/71	7 (8%)
	Ouerv: 38 TIT	VTADRTASSNDGY	ʹͲͶϨϾͲͰͲϷϜϒ	SI.PMTLRETPO	NSVSVTTSOON	はいしいよればいい	PAT. 95
55	T+	V TA + + Y	(+V+ T	+ MT R+IPO	2SV++++ Q+N	I DQ ++TL	+
	Sbjct: 43 TVI	VEGSATAPDDGENDY	SVTSTSAGTI	KMQMTQRDIP(	)SVTIVSQQRN	EDQQLQTLG	EVM 102
	Query: 96 LQA	TGTSRQIYGSDRAGY				ALADTGNAN	TAA 147
60	Shict: 103 FNT	G S+ SDRA Y LGISKSQADSDRALY		I NY ++GIP		AL+D	A -NT 154
00	SDJCC. TOS ERI	TGIONOGNOODNALI	ISKGr <u>V</u> .	IDNIMADGIE	TESKWILGE	NATODW	-AL 134
		VEVVRGVAGLPDGTO					
		VEVVRG GL GTO VEVVRGATGLMTGTO			+V AE G+ SDVSAEYGSWN	AD+ IKERYVADLQ	L SPL 214
65							
		GTLRGRLVSTFGRGI G +R R+V + I	DSWRQLERSRI DSW S	DAELYGILEYI GI++ I			ADA 266 +
		GKIRARIVGGYQNNI					
70	Ouerv: 267 PLS	YAVYDSQGYATAFGI	KDNPATNWS	NSRNRAT.NT.F	AGIEHRFNON	KLKARYDYT	RSR 326
. •	20023. 20. 200						250

	Sbjct:	275	+++ G + ++ + A +W+ + +F ++ +F W+ ++ WGGLPRWNTDGSSNSYDRARSTAPDWAYNDKEINKVFMTLKQQFADTWQATLNATHSEVE 334
5			FRQPYGVAGVLSIDHSTAATDLIPGYWHADPRTHSA-SMSLTGKYRLFG 374 F + Y A V D ++ PG+ W++ R A + G Y LFG
3	Sbjct:	335	F + Y A V D ++ PG+ W++ R A + G Y LFG FDSKMMYVDAYVNKADGMLVGPYSNYGPGFDYVGGTGWNSGKRKVDALDLFADGSYELFG 394
	Query:	375	REHDLIAGINGYKYASNKYGERSIIPNAIPNAYEFSRTGAYPQPSSFAQTIPQYDTRR 432 R+H+L+ G Y +N+Y +I P+ I + Y F+ G +PQ Q++ Q DT
10	Sbjct:	395	R+H+L+ G Y +N+Y +I P+ I + Y F+ G +PQ Q++ Q DT RQHNLMFG-GSYSKQNNRYFSSWANIFPDEIGSFYNFNGNFPQTDWSPQSLAQDDTTH 451
	Query:	433	QIGGYLATRFRAADNLSLILGGRYSRYRAGSYNSRTQGMTY-VSANRFTPYTGIVFDXXX 491 Y ATR AD L LILG RY+ +R + +TY + N TPY G+VFD
15	Sbjct:	452	MKSLYAATRVTLADPLHLILGARYTNWRVDTLTYSMEKNHTTPYAGLVFDIND 504
	Query:	492	XXXXXXXXXXFVPQLQKDEHGSYLKPVTGNNLEADIKGEWLEGRLNASAAVYRARKNNL 551 F PQ +D G YL P+TGNN E +K +W+ RL + A++R ++N+
	Sbjct:	505	NWSTYASYTSIFQPQNDRDSSGKYLAPITGNNYELGLKSDWMNSRLTTTLAIFRIEQDNV 564
20	Query:	552	ATAAGRDQSGNTYYRAANQAKTHGWEIEVGGRITPEWQIQAGYSQSKPRDQDGSRLN 608 A + G +G T Y+A + + G E E+ G IT WQ+ G ++ D +G+ +N
	Sbjct:	565	AQSTGTPIPGSNGETAYKAVDGTVSKGVEFELNGAITDNWQLTFGATRYIAEDNEGNAVN 624
25	Query:	609	PDSVPERSFKLFTAYHLAPEAPSGRTIGAGVRRQGETHTDPAALRIPNPAAKARAVANSR 668 P ++P + K+FT+Y L P P T+G GV Q +TD P RA
	Sbjct:	625	P-NLPRTTVKMFTSYRL-PVMPE-LTVGGGVNWQNRVYTDTVTPYGTFRA-+E 672
•	Query:	669	QKAYAVADIMARYRFNPRTELSLNVDNLFNKHYRTQPDRH-SYGALRTVNAAFTYRF 724 . Q +YA+ D+ RY+ L NV+NLF+K Y T + YG R + TY+F
30	Sbjct:	673	QGSYALVDLFTRYQVTKNFSLQGNVNNLFDKTYDTNVEGSIVYGTPRNFSITGTYQF 729

Based on this analysis, it was predicted that these proteins from N.meningitidis and N.gonorrhoede, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF23-1 (77.5kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 15A shows the results of affinity purification of the His-fusion protein, and Figure 15B shows the results of expression of the GST-fusion in *E.coli*. Purified His-fusion protein was used to immunise mice, whose sera were used for Western blot (Figure 15C) and for ELISA (positive result). These experiments confirm that ORF23-1 is a surface-exposed protein, and that it is a useful immunogen.

### 40 Example 80

35

The following partial DNA sequence was identified in N. meningitidis <SEQ ID 673>:

- 1 ATGCGCACGG CAGTGGTTTT GCTGTTGATC ATGCCGATGG CGGCTTCGTC
  51 GGCAATGATG CCGGAAATGG TGTGCGCGGG CGTGTCGCCG GGAACGGCAA
  101 TCATATCCAA GCCGACCGAA CAAACGGCGG TCATGGCTTC GAGTTTGTCC
  45 151 AGCGTCAGCA CGCCTGCTTC GGCGGCGGCA ATCATACCTT CGTCTTCGGA
  201 AACGGGGATA AACGCGCCAC TCAAACCCCC GACCGCGCTG GAAGCCATCA
  251 TGCCGCCTTT TTTCACGGCA TCGTTCAGCA ATGCCAAAGC TGCTGTTGTG
  301 CCGTGCGTAC CGCAGACGCT CAAGCCCATT TnTTCAAGAA TGCGTGCCAC
  351 TNAGTCGCCG ACGGGG..
- 50 This corresponds to the amino acid sequence <SEQ ID 674; ORF24>:
  - 1 MRTAVVLLLI MPMAASSAMM PEMVCAGVSP GTAIISKPTE QTAVMASSLS
  - 51 SVSTPASAAA IIPSSSETGI NAPLKPPTAL EAIMPPFFTA SFSNAKAAVV
  - 101 PCVPQTLKPI XSRMRATXSP TG...

# Further work revealed the complete nucleotide sequence <SEQ ID 675>:

	1	ATGCGCACGG	CAGTGGTTTT	GCTGTTGATC	ATGCCGATGG	CGGCTTCGTC
	51	GGCAATGATG	CCGGAAATGG	TGTGCGCGGG	CGTGTCGCCG	GGAACGGCAA
	101	TCATATCCAA	GCCGACCGAA	CAAACGGCGG	TCATGGCTTC	GAGTTTGTCC
5	151	AGCGTCAGCA	CGCCTGCTTC	GGCGGCGGCA	ATCATACCTT	CGTCTTCGGA
	201	AACGGGGATA	AACGCGCCAC	TCAAACCCCC	GACCGCGCTG	GAAGCCATCA
	251	TGCCGCCTTT	TTTCACGGCA	TCGTTCAGCA	ATGCCAAAGC	TGCTGTTGTG
	301	CCGTGCGTAC	CGCAGACGCT	CAAGCCCATT	TCTTCAAGAA	TGCGTGCCAC
	351	TGAGTCGCCG	ACGGCGGGG	TCGGCGCCAG	CGACAAGTCG	AGAATACCAA
10	401	ACGGGATATT	CAGCATTTTT	GAGGCTTCGC	GGCCGATGAG	TTCGCCCACG
	451	CGGGTAATTT	TGAAAGCAGT	TTTCTTCACT	ACTTCCGCAA	CTTCGGTCAA
	501	TGTCGTTGCA	TCTGAATTTT	CCAACGCGGC	TTTTACGACA	CCTGGGCCGG
	551	ATACGCCGAC	ATTGATAACG	GCATCCGCTT	CGCCCGAACC	ATGAAACGCG
	601	CCCGCCATAA	ACGGGTTGTC	TTCCACCGCG	TTGCAGAACA	CGACAATTTT
15	651	AGCGCAGCCG	AAACCTTCGG	GCGTGATTTC	CGCCGTGCGT	TTGACGGTTT
	701	CGCCCGCCAG	CTTGACCGCA	TCCATATTGA	TACCGGCACG	CGTACTGCCG
	751	ATATTGATGG	AGCTGCACAC	AATATCGGTA	GTCTTCATCG	CTTCGGGAAT
	801	GGAGCGGATT	AACACCTCAT	CCGAAGGCGA	CATCCCTTTT	TGCACCAACG
	851	CGGAAAAACC	GCCGATAAAA	GACACACCGA	TGGCTTTGGC	AGCTTTATCC
20	901	AAAGTTTGCG	CCACGCTGAC	GTAA		

This corresponds to the amino acid sequence <SEQ ID 676; ORF24-1>:

	1	MRTAVVLLLI	MPMAASSAMM	PEMVCAGVSP	GTAIISKPTE	<b>QTAVMASSLS</b>
	51	SVSTPASAAA	IIPSSSETGI	NAPLKPPTAL	EAIMPPFFTA	SFSNAKAAVV
	101	PCVPQTLKPI	SSRMRATESP	TAGVGASDKS	RIPNGIFSIF	EASRPMSSPT
25	151	RVILKAVFFT	TSATSVNVVA	SEFSNAAFTT	PGPDTPTLIT	ASASPEP*NA
	201	PAINGLSSTA	LQNTTILAQP	KPSGVISAVR	LTVSPASLTA	SILIPARVLP
	251	ILMELHTISV	VFIASGMERI	NTSSEGDIPF	CTNAEKPPIK	DTPMALAALS
	301	KVCATLT*				

Computer analysis of this amino acid sequence gave the following results:

# 30 Homology with a predicted ORF from N.meningitidis (strain A)

ORF24 shows 96.4% identity over a 307 aa overlap with an ORF (ORF24a) from strain A of N. meningitidis:

		10	20	30	40	50	60
	orf24a.pep	MRTAVVLLLIMPMA	assammpem	VCAGVSPGTAI	ISXPTEQTAV	'IASSLSNVST	PASAAA
35			111111111	1111111111	11   1111111	: ! ! ! ! ! : ! ! !	11111
	orf24	MRTAVVLLLIMPMA	ASSAMMPEM	VCAGVSPGTAI	ISKPTEQTAV	MASSLSSVST	PASAAA
		10	20	30	40	50	60
		70	80	90	100	110	120
40	orf24a.pep	IIPSSSXTGINAPL	KPPTALEAII	MPPFFTASFSN	IAKAAVVPCVE	QTLKPISSRM	RATESP
		111111 111111	111111111	11111111111	1111111111	111111111111111111111111111111111111111	111111
	orf24	IIPSSSETGINAPL	KPPTALEAI	MPPFFTASFSN	IAKAAVVPCVE	QTLKPISSRM	RATESP
		70	80	90	100	110	120
45		130	140	150	160	170	180
	orf24a.pep	TAGVGASDKSRIPN	GIFSIFEAS	RPMSSPTRVII	KAVFFTTSAT	SVNVVASEFS	NAAFTT
	• •		111111111			11111111111	111111
	orf24	TAGVGASDKSRIPN	GIFSIFEAS	RPMSSPTRVII	KAVEFTTSAT	SVNVVASEFS	NAAFTT
		130	140	150	160	170	180
50							
		190	200	210	220	230	240
	orf24a.pep	PGPDTPTLITASAS	PEPXNAPAI	XGLSSXALONT	TILAOPKPSS	SVISXVRLMVS	PASLTA
	• •		111111111	1111:111	HHÜHH:		111111
	orf24	PGPDTPTLITASAS	PEPXNAPAI	NGLSSTALON	TILAOPKPSO	VISAVRLTVS	PASLTA
55		190	200	210	220	230	240
		250	260	270	280	290	300
	orf24a.pep	SILIPARVLPILME					
60	orf24	SILIPARVLPILME		ASGMERINTS	SEGDIPFCTNA	EKPPIKDTPN	
••		250	260	270	280	290	300

# 5 The complete length ORF24a nucleotide sequence <SEQ ID 677> is:

	1	ATGCGCACGG	CAGTGGTTTT	GCTGTTGATC	ATGCCGATGG	CGGCTTCGTC
•	51	GGCAATGATG	CCGGAAATGG	TGTGCGCGGG	TGTGTCGCCG	GGAACGGCAA
•	101	TCATATCCAA	NCCGACCGAA	CAAACGGCGG	TCATCGCTTC	GAGTTTATCC
	151	AACGTCAGCA	CGCCTGCTTC	GGCGGCGCA	ATCATACCTT	CGTCTTCGGA
10	201	NACGGGGATA	AACGCGCCAC	TCAAACCGCC	AACCGCGCTC	GAAGCCATCA
•	251				ATGCCAAAGC	
	301	CCGTGCGTAC	CGCAGACGCT	CAAACCCATT	TCTTCAAGAA	TGCGCGCCAC
	351				CGACAAGTCG	
	401	ACGGGATATT	CAGCATTTTT	GAGGCTTCGC	GGCCGATGAG	TTCGCCCACG
15	451	CGGGTAATTT	TGAAGGCGGT	TTTCTTCACA	ACTTCGGCAA	CTTCGGTCAA
	501	TGTCGTTGCA	TCCGAATTTT	CCAACGCGGC	TTTTACGACA	CCCGGGCCGG
	551	ATACGCCGAC	ATTAATCACA	GCATCCGCTT	CGCCTGAGCC	GTGAAACGCG
•	601	CCCGCCATAN	ACGGGTTGTC	TTCCNCCGCG	TTGCAGAACA	CGACGATTTT
••	651				ANCCGTGCGT	
20	701	CGCCCGCCAG	TCTGACCGCG	TCCATATTGA	TACCGGCGCG	CGTACTGCCG
	751				GTCTTCATCG	
	801	GGAACGGATN	AACACCTCGT	CAGAAGGCGA	CATACCTTTT	TGCACCAGCG
	851	CGGAAAAGCC	GCCAATAAAA	GACACGCCGA	TGGCTTTGGC	AGCCTTATCC
	901	AAAGTTTGCG	CCACGCTGAC	GTAA	•	

# 25 This encodes a protein having amino acid sequence <SEQ ID 678>:

	1	MRTAVVLLLI	MPMAASSAMM	PEMVCAGVSP	GTAIISXPTE	OTAVIASSLS
	51	NVSTPASAAA	IIPSSSXTGI	NAPLKPPTAL	EAIMPPFFTA	SFSNAKAAVV
	101	PCVPQTLKPI	SSRMRATESP	<b>TAGVGASDKS</b>	RIPNGIFSIF	EASRPMSSPT
	151	RVILKAVFFT	TSATSVNVVA	SEFSNAAFTT	PGPDTPTLIT	ASASPEP*NA
30	201	PAIXGLSSXA	LQNTTILAQP	KPSSVISXVR	LMVSPASLTA	SILIPARVLP
	251	ILMELHTISV	VFIASGMERX	NTSSEGDIPF	CTSAEKPPIK	DTPMALAALS
		KVCATLT*				

It should be noted that this protein includes a stop codon at position 198.

# ORF24a and ORF24-1 show 96.4% identity in 307 aa overlap:

						•	
35		10	20	30	40	50	60
	orf24a.pep	MRTAVVLLLIMPMA	AASSAMMPEM		TISXPTEOTAL	TASSTSWS	מממפממי
	<b>-</b>	1111111111111				:           :	
	orf24-1	MRTAVVLLLIMPM	ADGSAMMDEM				
		10	20	30			
40	_	. 10	. 20	30	40	50	60
40		70	00				
			80	90	100	110	120
	orf24a.pep	IIPSSSXTGINAPI					
		111111 111111			11111111111		
45	orf24-1	IIPSSSETGINAPI	LKPPTALEAI		NAKAAVVPCVI	POTLKPISSR	RATESP
45		70	80	90	100	110	120
	-						
		130	140	150	160	170	180
	orf24a.pep	TAGVGASDKSRIPN	NGIFSIFEAS	RPMSSPTRVI	LKAVFFTTSAT	SVNVVASEFS	NAAFTT
			111111111	11111111111			
50	orf24-1	TAGVGASDKSRIP	GIFSIFEAS	RPMSSPTRVI	LKAVEFTTSAT	SVNVVASEES	ייייים ממווי
		130	140	150	160	170	180
•				200	100	170	100
		190	200	210	220	220	0.40
	orf24a.pep	PGPDTPTLITASAS				230	240
55	orresu.pep	111111111111111111111111111111111111111	DEFERNACAT				
	orf24-1				<u> </u>	111 111 11	
•	01124-1	PGPDTPTLITASAS	SPEPXNAPAL	NGLSSTALON			
		190	200	210	220	230	240
<b>60</b>		250	260	270	280	290	300
60	orf24a.pep	SILIPARVLPILM	ELHTISVVFI	<b>ASGMERXNTS</b>	SEGDIPFCTSA	EKPPIKDTP	MALAALS
		11313111111111	[	111111 111	111111111111111111111111111111111111111	THE HALL OF	111111
	orf24-1	SILIPARVLPILM	ELHTISVVFI	ASGMERINTS	SEGDIPFCTNA	EKPPIKOTP	ALAALS
		250	260	270	280	290	300

	orf24a.pep	KVCATLTX				
5	orf24-1	 KVCATLTX				
	Homology with a pr	edicted ORF from N.s	zonorrhoeae			
					dicted ORF (ORF24n	g) from
	N.gonorrhoeae:					
10	orf24.pep				TAVMASSLSSVSTPASAAA	60
	orf24ng					60
15	orf24.pep				PCVPQTLKPIXSRMRATXSP	120
15	orf24ng				CVPQTLKPISSRMRATESP	120
	orf24.pep	TG  :				122
20	orf24ng				SATSVRLTASEFSSAALTT	180
	The complete length	ORF24ng nucleotide	e sequence <	SEQ ID 679	> is:	
		CGCACGG CGGTGGTTTT				
25	101 TCA	TGTCCAA ACCAACGGAG	CAGACGGCGG	TCATGGCTTC	GAGTTTGTCC	
25		GTCAACA CGCCTGCCTC GGGGGATA AACGCGCCGC		-		
	251 TGC	CGCCCTT TTTCACGGCA	TCGTTCAGCA	ATGCCAAAGC	TGCTGTTGTG	
		TGCGTAC CGCAGACGCT				
30		GGATATT CAGCATTTTT				
		GTGATTT TGAAAGCGGT				
		GACCGCG TCCGAATTTT CGCCGAC ATTAATCACA				
		GCCATAA ACGGATTGTC				
35		GCAGCCG AAACCTTCGG				
		CTGCCAG CTTGACCGCA				
	801 · GG	ACGGATC AACACCTCAT	CCGAAGGCGA	CATACCTTTT	TGCACCAGCG	
40		SAAAAGCC GCCGATAAAG AGTCTGCG CCACGCTGAC		TGGCTTTGGC	TGCCTTGTCC	
+0				770 TO 600		
	This encodes a prote	ein having amino acid	sequence <	SEQ ID 680:	>:	
		AVVLLLI MPMAASSAMM				
		TPASAAA IIPSSSETGI POTLKPI SSRMRATESP				
45	151 RV	LKAVFFT TSATSVRLTA	SEFSSAALTT	PGPDTPTLIT	ASASPEPWNA	
		NGLSSTA LQNTTILAQP				
		MELHTISV VFIASGTERI CATLT*	NISSEGDIFF	CISAERPPIR	DIPMALAALS	
	ORF24ng and ORF	24-1 show 96.1% ider	ntity in 307 a	a overlap:		
50	_	10	20	20		50
30	orf24-1.pep	10 MRTAVVLLLIMPMAASS	20 SAMMPEMVCAGV		40 50 FEQTAVMASSLSSVSTPASA	60 AA
		111111111111111111		11111111111	111111111111111111111111111111111111111	11
	orf24ng	MRTAVVLLLIMPMAASS	SAMMPEMVCAGV 20		reqtavmasslssvntpasa 40 50	AA 60
55						
	arf21=1 na-	70	80 מספאד גים ז גיייכ			20
	orf24-1.pep				VVPCVPQTLKPISSRMRATE 	
60	orf24ng				VVPCVPQTLKPISSRMRATE	
60		70	80	90 10	00 110 1	120

-390-

		130	140	150	160	170	180
	orf24-1.pep	TAGVGASDKSRIPNGI	FSIFEASRPM	SSPTRVILKA	VFFTTSATSV	NVVASEFSNA	AFTT
		11111111111111111	1111111111	111111111	1111111111	::!!!!!:!	
_	orf24ng	TAGVGASDKSRMPNGI					
<b>.</b>		130	140	150	160	170	180
	•	190	200	210	220	230	240
	orf24-1.pep	PGPDTPTLITASASPE					
	Oller Tipop		1 IIIIIIII		TUĞEKESGAT		IIII
10	orf24ng	PGPDTPTLITASASPE	יוווווווווו סאדאס מאשםי	ווווווווווו דייילות המייטר	11111111111 1100KBCCUT		
10	VIII-1119	190	200	210	220	230	240
		100	200	210	220	230	240
		250	260	270	280	290	300
	orf24-1.pep	SILIPARVLPILMELH	TISVVFIASG	MERINTSSEG	DIPFCTNAEK	PPIKDTPMAL	AALS
15	•	- 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1111111111	111111111	1111111:111	1111111111	1111
	orf24ng	SILIPARVLPILMELH	TISVVFIASG	TERINTSSEG	DIPFCTSAEK	PPIKDTPMAL	AALS
		250	260	270	280	290	300
20	orf24-1.pep	KVCATLTX			•		
		1111111			•		
•	orf24ng	KVCATLTX					
	· · ·						

Based on this analysis, including the presence of a putative leader sequence (first 18 aa – double-underlined) and putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 81

**25** 

The following partial DNA sequence was identified in N.meningitidis <SEQ ID 681>:

```
1 .ACCGACGTGC AAAAAGAGTT GGTCGGCGAA CAACGCAAGT GGGCGCAGGA
51 AAAAATCAGC AACTGCCGAC AAGCCGCCGC GCAGGCAGCA CGCCAGGAAT
101 ACGCCGAATA CCTCAAGCTG CAATGCGACA CGCGGATGAC GCGCGAACGG
151 ATACAGTATC TTCGCGGCTA TTCCATCGAT TAG
```

This corresponds to the amino acid sequence <SEQ ID 682; ORF25>:

1 ..TDVQKELVGE QRKWAQEKIS NCRQAAAQAD RQEYAEYLKL QCDTRMTRER
51 IQYLRGYSID \*

Further work revealed the complete nucleotide sequence <SEQ ID 683>:

	1 -	ATGTATCGGA	AACTCATTGC	GCTGCCGTTT	GCCCTGCTGC	TTGCCGCTTG
	51	CGGCAGGGAA	GAACCGCCCA	AGGCATTGGA	ATGCGCCAAC	CCCGCCGTGT
40	101	TGCAAGGCAT	ACGCGGCAAT	ATTCAGGAAA	CGCTCACGCA	GGAAGCGCGT
	151	TCTTTCGCGC	GCGAAGACGG	CAGGCAGTTT	GTCGATGCCG	ACAAAATTAT
	201	CGCCGCCGCC	TACGGTTTGG	CGTTTTCTTT	GGAACACGCT	TCGGAAACGC
	251	AGGAAGGCGG	GCGCACGTTC	TGTATCGCCG	ATTTGAACAT	TACCGTGCCG
	301	TCTGAAACGC	TTGCCGATGC	CAAGGCAAAC	AGCCCCCTGT	TGTACGGGGA
45	351	AACTGCTTTG	TCGGATATTG	TGCGGCAGAA	GACGGGCGGC	AATGTCGAGT
. '	401	TTAAAGACGG	CGTATTGACG	GCAGCCGTCC	GCTTCCTGCC	CGTCAAAGAC
	451	GGTCAGACGG	CATTTGTCGA	CAACACGGTC	GGTATGGCGG	CGCAAACGCT
	501	GTCTGCCGCG	CTGCTGCCTT	ACGGCGTGAA	GAGCATCGTG	ATGATAGACG
	551	GCAAGGCGGT	GAAAAAAGAA	GACGCGGTCA	GGATTTTGAG	CGGAAAAGCC
50	601	CGTGAAGAAG	AACCGTCCAA	ACCCACGCCC	GAAGACATTT	TGGAACACAA
	651	TGCCGCCGGC	GGCGATGCGG	GCGTACCCCA	AGCCGCAGAA	GGCGCGCCCG
	701	AACCGGAAAT	CCTGCATCCT	GACGACGGCG	AGCGTGCCGA	TACCGTTACC
	751	GTATCACGGG	GCGAAGTGGA	AGAGGCGCGC	GTACAAAACC	AGCGTGCGGA
	801	ATCCGAAATT	ACCAAACTTT	GGGGAGGACT	CGATACCGAC	GTGCAAAAAG
55	851	AGTTGGTCGG	CGAACAACGC		AGGAAAAAAT	CAGCAACTGC
	901	CGACAAGCCG	CCGCGCAGGC	AGACCGGCAG	GAATACGCCG	AATACCTCAA
	951		· · · · · · · ·		ACGGATACAG	
	1001	GCTATTCCAT				
				•		

This corresponds to the amino acid sequence <SEQ ID 684; ORF25-1>:

```
1 MYRKLIALPF ALLLAACGRE EPPKALECAN PAVLQGIRGN IQETLTQEAR
51 SFAREDGRQF VDADKIIAAA YGLAFSLEHA SETQEGGRTF CIADLNITVP
101 SETLADAKAN SPLLYGETAL SDIVRQKTGG NVEFKDGVLT AAVRFLPVKD
151 GQTAFVDNTV GMAAQTLSAA LLPYGVKSIV MIDGKAVKKE DAVRILSGKA
201 REEEPSKPTP EDILEHNAAG GDAGVPQAAE GAPEPEILHP DDGERADTVT
251 VSRGEVEEAR VQNQRAESEI TKLWGGLDTD VQKELVGEQR KWAQEKISNC
301 RQAAAQADRQ EYAEYLKLQC DTRMTRERIQ YLRGYSID*
```

Computer analysis of this amino acid sequence gave the following results:

# 10 Homology with a predicted ORF from N. meningitidis (strain A)

ORF25 shows 98.3% identity over a 60aa overlap with an ORF (ORF25a) from strain A of N. meningitidis:

```
20
                                                     10
                                              TDVQKELVGEQRKWAQEKISNCRQAAAQAD
         orf25.pep
                                               иншин ишиншіній
15
                    VTVSRGEVEEARVQNQRAESEITKLWGGLDTDVQKELVGEXRKWAQEKISNCRQAAAQAD
         orf25a
                                                               300
                            260
                                     270
                                              280
                                                      290
                    250
                            40
                     ROEYAEYLKLOCDTRMTRERIQYLRGYSIDX
20
         orf25.pep
                     RQEYAEYLKLQCDTRMTRERIQYLRGYSIDX
         orf25a
                                     330
```

The complete length ORF25a nucleotide sequence <SEQ ID 685> is:

25	1	ATGTATCGGA	AACTCATTGC	GCTGCCGTTT	GCCCTGCTGC	TTGCCGCTTG
23	51	CGGCAGGGAA	GAACCGCCCA	AGGCATTGGA	ATGCGCCAAC	CCCGCCGTGT
	101	TGCAANGCAT	ACGCNGCAAT	ATTCAGGAAA	CGCTCACGCA	GGAAGCGCGT
	151	TCTTTCGCGC	GCGAAGACNG	CANGCAGTTT	GTCGATGCCG	ACNAAATTAT
	201	CGCCGCCGCC	TANGNTNNGN	NGNTNTCTTT	GGAACACGCT	TCGGAAACGC
30	251	AGGAAGGCGG			ATTTGAACAT	TACCGTGCCG
50	301	TCTGAAACGC	TTGCCGATGC	CAAGGCAAAC		TGTACGGGGA
	351	AACCGCTTTG	TCGGATATTG	TGCGGCAGAA		
	401	TTAAAGACGG	CGTATTGACG	GCAGCCGTCC	GCTTCCTACC	CGTCAAAGAC
	451	GGTCAGANGG	CATTTGTCGA	CAACACGGTC	GGTATGGCGG	
35	501	GTCTGCCGCG	TTGCTGCCTT	ACGGCGTGAA	GAGCATCGTG	ATGATAGACG
<i>33</i>	551	GCAAGGCGGT	AAAAAAAGAA	GACGCGGTCA	GGATTNTGAG	CNGANAAGCC
	601	CGTGAANAAG		ANCCNNGCCC		TGGAACATAA
	651	TGCCGCCGGA	GGGGATGCAG	ACGTACCCCA	AGCCGGAGAA	
	701	AACCGGAAAT	CCTGCATCCT	GACGACGGCG	AGCGTGCCGA	
40	751	GTATCACGGG	GCGAAGTGGA	AGAGGCGCGN	GTACAAAACC	
40	801	ATCCGAAATT	ACCAAACTTT	GGGGAGGACT	CGATACCGAC	
	851	AGTTGGTCGG	CGAANAACGC	AAGTGGGCGC	AGGAAAAAAT	
	901	CGACAAGCCG	CCGCGCAGGC	AGACCGGCAG	GAATACGCCG	
	951	GCTGCAATGC	GACACGCGGA	TGACGCGCGA	ACGGATACAG	TATCTTCGCG
45	1001	GCTATTCCAT	CGATTAG			

This encodes a protein having amino acid sequence <SEQ ID 686>:

```
1 MYRKLIALPF ALLLAACGRE EPPKALECAN PAVLQXIRXN IQETLTQEAR
51 SFAREDXXQF VDADXIIAAA XXXXXSLEHA SETQEGGRTF CXADLNITVP
101 SETLADAKAN SPLLYGETAL SDIVRQKTGG NVEFKDGVLT AAVRFLPVKD
151 GQXAFVDNTV GMAAQTLSAA LLPYGVKSIV MIDGKAVKKE DAVRIXSXXA
201 REXEPSKXXP EDILEHNAAG GDADVPQAGE DAPEPEILHP DDGERADTVT
251 VSRGEVEEAR VQNQRAESEI TKLWGGLDTD VQKELVGEXR KWAQEKISNC
301 RQAAAQADRQ EYAEYLKLQC DTRMTRERIQ YLRGYSID*
```

ORF25a and ORF25-1 show 93.5% identity in 338 aa overlap:

55		10 20 30 40 50 60						
	orf25a.pep	MYRKLIALPFALLL	111111111	111111111111	]	1111111111	11 11	
	orf25-1	MYRKLIALPFALLL	AACGREEPP	KALECANPAVL	QGIRGNIQET	LTQEARSFAR	EDGRQF	

-392-

		10	20 .	30	40	50	60
		70	80	90	100	110	120
_	orf25a.pep	VDADXIIAAAXXXX	XSLEHASETQ	EGGRTFCXAL	LNITVPSETL	ADAKANSPLI	
<b>.</b> 5		1111 11111	_		шшш		
	orf25-1	VDADKIIAAAYGLA					
		70	80	90	100	110	120
		130	140	150	160	170	180
10	orf25a.pep	SDIVRQKTGGNVE	KDGVLTAAVR	.FLPVKDGQX <i>I</i>	FVDNTVGMAA	QTLSAALLPY	GVKSIV
	orf25-1	SDIVRQKTGGNVE					
		130	140	150	160	170	180
15		190	200	210	220	230	240
	orf25a.pep	MIDGKAVKKEDAVE					240 סעודים
	orf25-1	MIDGKAVKKEDAVE	RILSGKAREEE				
		190	200	210	220	230	240
20			•		•		
		250	260	270	280	290	300
	orf25a.pep	DDGERADTVTVSRC	EVEEARVONC	RAESEITKL	<b>IGGLDTDVQKE</b>	LVGEXRKWAC	EKISNC
	orf25-1			111111111			
25	01125-1	DDGERADTVTVSRG 250	SEVEEARVONO 260	RAESELTKLV 270	VGGLDTDVQKE 280		
25		230	200	210	260	290	300
•		310	320	330	339	į	
	orf25a.pep	RQAAAQADRQEYAE				•	
			ШПП	1111111111	11111		
30	orf25-1	RQAAAQADRQEYAE			SYSIDX		
		310	320	330	•		

## Homology with a predicted ORF from N.gonorrhoeae

ORF25 shows 100% identity over a 60aa overlap with a predicted ORF (ORF25ng) from

## 35 N.gonorrhoeae:

# The complete length ORF25ng nucleotide sequence <SEQ ID 687> is:

1	<b>ATGTATCGGA</b>	AACTCATTGC	GCTGCCGTTT	GCCCTGCTGC	TTGCAGCGTG
51	CGGCAGGGAA	GAACCGCCCA	AGGCGTTGGA	ATGCGCCAAC	CCCGCCGTGT
101	TGCAGGACAT	ACGCGGCAGT	ATTCAGGAAA	CGCTCACGCA	GGAAGCGCGT
151					
201	CGCCGCCGCC	TACGGTTTGG	CGTTTTCTTT	GGAACACGCT	TCGGAAACGC
251	AGGAAGGCGG	GCGCACGTTC	TGTATCGCCG	ATTTGAACAT	TACCGTGCCG
301	TCTGAAACGC	TTGCCGATGC	CGAGGCAAAC	AGCCCCCTGC	TGTATGGGGA
351	AACGTCTTTG	GCAGACATCG	TGCAGCAGAA	GACGGGCGGC	AATGTCGAGT
401	TTAAAGACGG	CGTATTGACG	GCAGCCGTCC	GCTTCCTGCC	CGCCAAAGAC
451	GCTCGGACGG	CATTTATCGA	CAACACGGTC	GGTATGGCGA	CGCAAACGCT
	GTCTGCCGCG	TTGCTGCCTT	ACGGCGTGAA	GAGCATCGTG	ATGATAGACG
	GCAAGGCGGT	GACAAAAGAA	GACGCGGTCA	GGGTTTTGAG	CGGCAAAGCC
601	CGTGAAGAAG	AACCGTCCAA	ACCCACCCC	GAAGACATTT	TGGAACACAA
651	TGCCGCCGGC	GGCGATGCGG	GCGTACCCCA	AGCCGCAGAA	GGCGCACCCG
701	AACCCGAAAT	CCTGCATCCC	GACGACGTCG	AGCGTGCCGA	TACCGTTACC
751	GTATCACGGG	GCGAAGTGGA	AGAGGCGCGC	GTACAAAACC	AACGTGCGGA
	AGTTGGTCGG	CGAACAGCGC	AAGTGGGCGC	AGGAAAAAAT	CAGcaactgc
	cgacaagccg	CCGCGCAGGC	AGACCGGCAG	GAATACGCCG	AATACCTCAA
			TGACGCGCGA	ACggaTACAG	TATCTTCGCG
1001	GCTATTCCAT	CGATTAG			
	51 101 151 201 251 301 351 401 451 501 551 601 651 701	51 CGGCAGGAA 101 TGCAGGACAT 151 TCTTTCGCGC 201 CGCCGCCGCC 251 AGGAAGCCG 301 TCTGAAACGC 351 AACGTCTTTG 401 TTAAAGACGG 451 GCTCGGACGG 501 GTCTGCCGCG 551 GCAAGGCGGT 601 CGTGAAGAAG 651 TGCCGCCGGC 701 AACCCGAAAT 751 GTATCACGGG 801 ATCCGAAATT 851 AGTTGGTCGG 901 CGACAAGCCG 951 GCTCCAATGC	51 CGGCAGGAA GAACCGCCCA 101 TGCAGGACAT ACGCGGCAGT 151 TCTTTCGCGC GCGAAGACGG 201 CGCCGCCGCC TACGGTTTGG 251 AGGAAGGCGG GCGCACGTTC 301 TCTGAAACGC TTGCCGATGC 351 AACGTCTTTG GCAGACATCG 401 TTAAAGACGG CGTATTGACG 451 GCTCGGACGG CATTTATCGA 501 GTCTGCCGCG TTGCTGCCTT 551 GCAAGGCGGT GACAAAAGAA 601 CGTGAAGAA AACCGTCCAA 651 TGCCGCCGGC GGCGATGCGG 701 AACCCGAAAT CCTGCATCCC 751 GTATCACGG GCCAAGTGGA 801 ATCCGAAAT ACCAAACTTT 851 AGTTGGTCGG CGAACAGCGC 901 CGACAAGCCG CGCGCAGGC	51 CGGCAGGGAA GAACCGCCCA AGGCGTTGGA 101 TGCAGGACAT ACGCGGCAGT ATTCAGGAAA 151 TCTTTCGCGC GCGAAGACGG CAGGCAGTTT 201 CGCCGCCGCC TACGGTTTGG CGTTTTCTTT 251 AGGAAGGCGG GCGCACGTTC TGTATCGCCG 301 TCTGAAACGC TTGCCGATGC CGAGGCAAAC 351 AACGTCTTTG GCAGACATCG TGCAGCAGAA 401 TTAAAGACGG CGTATTGACG GCAGCCGTCC 451 GCTCGGACGG CATTTATCGA CAACACGGTC 501 GTCTGCCGCG TTGCTGCCTT ACGGCGTGAA 551 GCAAGGCGGT GACAAAAGAA GACGCGGTCA 601 CGTGAAGAAG AACCGTCCAA ACCCACCCCC 651 TGCCGCCGC GGCGATGCGG GCGTACCCCA 651 TGCCGCCGGC GGCGATGCGG GCGTACCCCA 701 AACCCGAAAT CCTGCATCCC GACGACGTCG 751 GTATCACGGG GCGAAGCGG AGAGGGCGCG 801 ATCCGAAATT ACCAAACTTT GGGGAGGGCCGC 801 ATCCGAAATT ACCAAACTTT GGGGAGGGACT 851 AGTTGGTCGG CGAACAGCGC AAGTGGGCGC 901 CGACAAGCCG CCGCGCAGGC AGACCGCCAA	CGGCAGGGAA GAACCGCCCA AGGCGTTGGA ATGCGCCAAC TGCAGGACAT ACGCGGCAGT ATTCAGGAAA CGCTCACGCA TCTTTCGCGC GCGAAGACGG CAGGCAGTTT GTCGATGCCG CGCCGCCGCC TACGGTTTGG CGTTTTCTTT GCAACACGCT AGGAAGGCGG GCGCACGTTC TGTATCGCCG ATTTGAACAT TCTGAAACGC TTGCCGATGC CGAGGCAAAC AGCCCCCTGC AACGTCTTTG GCAGACATCG TGCAGCAGAA GACGGCGGCG TTAAAGACGG CGTATTGACG GCAGCCGTCC GCTTCCTGCC CGTCGACGG CATTTATCGA CAACACGGTC GGTATGGCGA GCCAGGCGGT GACAAAAAGAA GACGGCGCAGAA GACGACTCTG GCAAGGCGGT GACAAAAAGAA GACGGCTCA GGGTTTTGAG CGTGAAGAAG AACCGTCCAA ACCCACCCC GAAGACATTT TGCCGCCGG GGCGATGCGG GCGTACCCCA AGCCGCAGAA TOI AACCCGAAAT CCTGCATCCC GACGACGTCC AGCGTGCCGA TOI AACCCGAAAT ACCAACCCC GACGACGCG TTATCACGGG GCGAAGTGGA AGAGGCGCGC GTACAAAACC TTATCACGGG GCGAAGTGGA AGAGGCGCGC GTACAAAACC CTGCAAATT ACCAAACTTT GGGGAGGACT CGATACCGAC CGAACAGCCG CGAACAGCGC AAGTGGGCGC AGGAAAAAAAT CCGAAATT ACCAAACTTT GGGGAGGACT CGATACCGAC CGAACAGCCG CGAACAGCGC AAGTGGGCGC AGGAAAAAAAT CCGAAATT ACCAAACTTT GGGGAGGACT CGATACCGAC CGAACAGCCG CGAACAGCGC AAGTGGGCGC AGGAAAAAAAT CCGAAATGCGCC CGAACAGCGC AAGTGGGCGC AGGAAAAAAAT CCGACAGCGC CGCGCAGGC AGACCGCCAG GAATACGCCG CGAACAGCCG CGAACAGCGC AAGTGGGCGCA ACGGAAAAAAAT CCGCCCAAGCCGC AGGAAAAAAAAT CCGACAGCGC CGCCCCAGGC AGACCGCCAG ACGGAAAAAAAT CCGCCCAGGC CGCCCCGCCAGGC AGGCCCGCAGAAAAAAAA

65 This encodes a protein having amino acid sequence <SEQ ID 688>:

50

55

5	51 101 151 201	SFAREDGRQF SETLADAEAN ARTAFIDNTV REEEPSKPTP	VDADKIIAAA SPLLYGETSL GMATQTLSAA EDILEHNAAG	YGLAFSLEHA ADIVQQKTGG LLPYGVKSIV GDAGVPQAAE	PAVLQDIRGS SETQEGGRTF NVEFKDGVLT MIDGKAVTKE GAPEPEILHP VQKELVGEQR	CIADLNITVP AAVRFLPAKD DAVRVLSGKA DDVERADTVT
	251	VSRGEVEEAR	VQNQRAESEI		VQKELVGEQR	

# ORF25ng and ORF25-1 show 95.9% identity in 338 aa overlap:

10	orf25-1.pep orf25ng	10 MYRKLIALPFALLLA !!!!!!!!!!!!! MYRKLIALPFALLLA 10	11111111	11111111111	î malin	HÜHHH	111111
15		70	80	90	100	110	120
	orf25-1.pep	VDADKIIAAAYGLAF	SLEHASET	QEGGRTFCIAD		ADAKANSPLI	
	orf25ng	VDADKIIAAAYGLAF	SLEHASET	QEGGRTFCIAD	LNITVPSETI	ADAEANSPLI	YGETSL
20		70	80	90	100	110	120
20		130	140	150	160	170	180
	orf25-1.pep	SDIVROKTGGNVEFK					
	orf25ng	:       :		:  ::   			
25	ortzang	130	140	150	160	170	180
		190	200	210	220	230	240
	orf25-1.pep	MIDGKAVKKEDAVRI					
•		1111111 1111111:	:			NI N	
30	orf25ng	MIDGKAVTKEDAVRV	/LSGKAREE 200	EPSKPTPEDII 210	EHNAAGGDAU 220	230	240
	orf25-1.pep	250 DDGERADTVTVSRGE	260	270 ODDESETTKIJ	280 IGGI DYDVOKI	290	300 DEKTSNC
35	01125-1.beb	11 111111111111	$\Pi\Pi\Pi\Pi\Pi$	ĨIIIIIIIII	111111111		1111111
	orf25ng	DDVERADTVTVSRGI			VGGLDTDVQKI 280	ELVGEQRKWA 290	QEKISNC 300
		250	260	270	280	290	300
		310	320	330	339		
40	orf25-1.pep	RQAAAQADRQEYAE					
	orf25ng	RQAAAQADRQEYAE					
	-	310	320	330			

Based on this analysis, including the presence of a predicted prokaryotic membrane lipoprotein lipid attchment site (underlined) in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF25-1 (37kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 16A shows the results of affinity purification of the GST-fusion protein, and Figure 16B shows the results of expression of the His-fusion in *E.coli*. Purified His-fusion protein was used to immunise mice, whose sera were used for Western blot (Figure 16C), ELISA (positive result), and FACS analysis (Figure 16D). These experiments confirm that ORF25-1 is a surface-exposed protein, and that it is a useful immunogen.

Figure 16E shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF25-1.

## Example 82

1251

1301

The following partial DNA sequence was identified in N.meningitidis <SEQ ID 689>

```
1 ATGCAGCTGA TCGACTATTC ACATTCATTT TTCTCGGTTG TGCCACCCTT
                     TTTGGCACTG GCACTTGCCG TCATTACCCG CCGCGTACTG CTGTCTTTAG
GCATCGGTAT TCTGGWySGC GTTGCCTTTT TGGTCGGCGG CAACCCCGTC
GACGGTCTGA CACACCTGAA AGACATGGTC GTCGGCTTGG CTTGGTCAGA
5
                101
                     CGsyGATTGG TCGCTGGGCA AACCAAAAAT CTTGGTTTTC CKGATACTTT
                     TGGGTATTTT TACTTCCCTG CTGACCTACT CCGGCAGCAA T.....
10
                                                   //
                                                        .....AC TTCGCTGGTA
                901
                     TTCGGCGGCA CTTGCGGCGT CTTTGCCGTC GTTCTCTGCA CGCTCGGCAC
                     GATTAAAACC GCCGACTATC CCAAAGCCGT TTGGCAGGGT GCGAAATCTA
                951
               1001
                     TGTTCGGCGC AATCGCCATT TTAATCCTCG CTTGGCTCAT CAGTACGGTT
15
                     GTCGGCGAAA TGCACACCGG CGATTACCTC TCCACACTGG TTGCGGGCAA
               1051
                     CATCCATCCC GGCTTCCTGC CCGTCATCCT CTTCCTGCTC GCCAGCGTGA
               1101
                     TGGCGTTTGC CACAGGCACA AGCTGGGGGA CGTTCGGCAT TATGCTGCCG
               1151
               1201
                     ATTGCCGCCG CCATGGCGGT CAAAGTCGAA CCCGCGCTGA TTATCCCGTG
                     TATGTCCGCA GTAATGGCGG GGGCGGTATG CGGCGACCAC TGCTCGCCCA
               1251
20
               1301
                     TTTCCGACAC GACCATCCTG TCGTCCACCG GCGCGCGCTG CAACCACATC
               1351
                     GACCACGTTA CCTCGCAACT GCCTTACGCC TTAACCGTTG CCGCCGCCGC
               1401
                     CGCATCGGGC TACCTCGCAT TGGGTCTGAC AAAATCCGCG CTGTTGGGCT
               1451
                     TTGGCACGAC AGGCATTGTA TTGGCGGTGC TGATTTTTCT GTTGAAAGAT
               1501
                     AAAAAA..
25
     This corresponds to the amino acid sequence <SEQ ID 690; ORF26>:
                     MQLIDYSHSF FSVVPPFLAL ALAVITRRVL LSLGIGILXX VAFLVGGNPV
                     DGLTHLKDMV VGLAWSDXDW SLGKPKILVF XILLGIFTSL LTYSGSN...
                251
30
                     FGGTCGVFAV VLCTLGTIKT ADYPKAVWQG AKSMFGAIAI LILAWLISTV
                     VGEMHTGDYL STLVAGNIHP GFLPVILFLL ASVMAFATGT SWGTFGIMLP
                     IAAAMAVKVE PALIIPCMSA VMAGAVCGDH CSPISDTTIL SSTGARCNHI
                     DHVTSQLPYA LTVAAAAASG YLALGLTKSA LLGFGTTGIV LAVLIFLLKD
                451
35
     Further work revealed the complete nucleotide sequence <SEO ID 691>:
                     ATGCAGCTGA TCGACTATTC ACATTCATTT TTCTCGGTTG TGCCACCCTT
                 51
                     TTTGGCACTG GCACTTGCCG TCATTACCCG CCGCGTACTG CTGTCTTTAG
                     GCATCGGTAT TCTGGTCGGC GTTGCCTTTT TGGTCGGCGG CAACCCCGTC
                     GACGGTCTGA CACACCTGAA AGACATGGTC GTCGGCTTGG CTTGGTCAGA
40
                     CGGCGATTGG TCGCTGGGCA AACCAAAAAT CTTGGTTTTC CTGATACTTT
                     TGGGTATTTT TACTTCCCTG CTGACCTACT CCGGCAGCAA TCAGGCGTTT
                251
                     351
                     GACCGCCTGC CTCGTGTTCG TAACCTTTAT CGACGACTAT TTCCACAGTC
                     TCGCCGTCGG TGCGATTGCC CGCCCCGTTA CCGACAAGTT TAAAGTTTCC
                401
45
                     CGCACCAAAC TCGCCTACAT CCTCGACTCC ACTGCCGCTC CTATGTGCGT
                451
                     GCTGATGCCC GTTTCAAGCT GGGGCGCGTC GATTATCGCC ACGCTTGCCG
                551
                     GACTGCTCGT TACCTACAAA ATCACCGAAT ACACGCCGAT GGGGACGTTT
                     GTCGCCATGA GCCTGATGAA CTATTACGCA CTGTTTGCCC TGATTATGGT
                     GTTCGTCGTC GCATGGTTTT CCTTCGACAT CGGCTCGATG GCACGTTTCG
                651
50
                     AACAAGCCGC GTTGAACGAA GCCCACGATG AAACTGCCGT TTCAGACGCT
                     ACCAAAGGTC GTGTTTACGC ACTGATTATT CCCGTTTTGG CCTTAATCGC
                751
                     CTCAACGGTT TCCGCCATGA TCTACACCGG CGCGCAGGCA AGCGAAACCT
                801
                851
                     TCAGCATTTT GGGGGCATTT GAAAACACGG ACGTAAACAC TTCGCTGGTA
                     TTCGGCGGCA CTTGCGGCGT CCTTGCCGTC GTTCTCTGCA CGCTCGGCAC
                901
55
                     GATTAAAACC GCCGACTATC CCAAAGCCGT TTGGCAGGGT GCGAAATCTA
                951
               1001
                     TGTTCGGCGC AATCGCCATT TTAATCCTCG CTTGGCTCAT CAGTACGGTT
               1051
                     GTCGGCGAAA TGCACACCGG CGATTACCTC TCCACACTGG TTGCGGGCAA
                     CATCCATCCC GGCTTCCTGC CCGTCATCCT CTTCCTGCTC GCCAGCGTGA
               1101
               1151
                     TGGCGTTTGC CACAGGCACA AGCTGGGGGA CGTTCGGCAT TATGCTGCCG
60
                     ATTGCCGCCG CCATGGCGGT CAAAGTCGAA CCCGCGCTGA TTATCCCGTG
               1201
```

TATGTCCGCA GTAATGGCGG GGGCGGTATG CGGCGACCAC TGCTCGCCCA

TTTCCGACAC GACCATCCTG TCGTCCACCG GCGCGCGCTG CAACCACATC

5 This corresponds to the amino acid sequence <SEQ ID 692; ORF26-1>:

						***************************************
	1	MOLIDYSHSF	FSVVPPFLAL	<u>ALAVITRRVL</u>	LSLGIGILIVG	VAFLVGGNPV
	51	DGLTHLKDMV	VGLAWSDGDW	SLGKPKILVF	LILLGIFTSL	<u>LTY</u> SGSNQAF
	101	ADWAKRHIKN	RRGAKMLTAC	LVFVTFIDDY	<b>FHSLAVGAIA</b>	RPVTDKFKVS
	151	RTKLAYILDS	TAAPMCVLMP	VSSWGASIIA	TLAGLLVTYK	ITEYTPMGTF
10	201	VAMSLMNYYA	LFALIMVFVV	AWFSFDIGSM	ARFEQAALNE	AHDETAVSDA
••	251	TKGRVYALII	PVLALIASTV	SAMIYTGAQA	SETFSILGAF	ENTDVNTS <u>LV</u>
	301	FGGTCGVLAV	VLCTLGTIKT	ADYPKAVWQG	AKSM <u>FGAIAI</u>	LILAWLISTV
	351	VGEMHTGDYL	STLVAGNIHP	GFLPVILFLL	<b>ASVMAFATGT</b>	SWGTFGIMLP
	401	<b>IAAAMAVKVE</b>	PALIIPCMSA	VMAGAVCGDH	CSPISDTTIL	SSTGARCNHI
15	451	DHVTSQLPYA	LTVAAAAASG	YLALGLTKSA	LLGFGTTGIV	LAVLIFLLKD
	501	KKRANA*				•

Computer analysis of this amino acid sequence gave the following results:

Homology with the hypothetical transmembrane protein HI1586 of *H.influenzae* (accession number P44263) ORF26 and HI1586 show 53% and 49% amino acid identity in 97 and 221 aa overlap at the

20 N-terminus and C-terminus, respectively:

```
MQLIDYSHSFFSVVPPFLALALAVITRRVXXXXXXXXXXXXVAFLVGGNPVDGLTHLKDMV 60
          Orf26 1
                    M+LID+S S +S+VP LA+ LA+ TRRV
          HI1586 14 MELIDFSSSVWSIVPALLAIILAIATRRVLVSLSAGIIIGSLMLSDWQIGSAFNYLVKNV 73
          Orf26 61 VGLAWSDXDWSLGKPKILVFXILLGIFTSLLTYSGSN 97
25
                    V L ++D + + I++F +LLG+ T+LLT SGSN
          HI1586 74 VSLVYADGEIN-SNMNIVLFLLLLGVLTALLTVSGSN 109
30
          Orf26 86 IFTSLLTYSGS--NTSLVFGGTCGVFAVVLCTL--GTIKTADYPKAVWQGAKSMFGXXXX 141
                    +F+ L T+ + TSLV GG C + L + + +Y ++ G KSM G
          HI1586 299 VFSVLGTFENTVVGTSLVVGGFCSIIISTLLIILDRQVSVPEYVRSWIVGIKSMSGAIAI 358
          Orf26 142 XXXXXXXSTVVGEMHTGDYLSTLVAGNIHPGFLPVILFLLASVMAFATGTSWGTFGIMLP 201
35
                           + +VG+M TG YLS+LV+GNI FLPVILF+L + MAF+TGTSWGTFGIMLP
          HI1586 359 LFFAWTINKIVGDMQTGKYLSSLVSGNIPMQFLPVILFVLGAAMAFSTGTSWGTFGIMLP 418
          Orf26 202 IAAAMAVKVEPALIIPCMSAVMAGAVCGDHCSPISDTTILSSTGARCNHIDHVTSQXXXX 261
                    IAAAMA P L++PC+SAVMAGAVCGDHCSP+SDTTILSSTGA+CNHIDHVT+Q
40
          HI1586 419 IAAAMAANAAPELLLPCLSAVMAGAVCGDHCSPVSDTTILSSTGAKCNHIDHVTTQLPYA 478
          S L GF T + L V+IF +K +
          HI1586 479 ATVATATSIGYIVVGFTYSGLAGFAATAVSLIVIIFAVKKR 519
45
```

Homology with a predicted ORF from N. meningitidis (strain A)

ORF26 shows 58.2% identity over a 502aa overlap with an ORF (ORF26a) from strain A of N. meningitidis:

50		10	20	30	40	50	60
	orf26.pep	MQLIDYSHSFFSVVI	PFLALALAV	/ITRRVLLSLO	IGILXXVAFL	<u>V</u> GGNPVDGLT	HLKDMV
		11111111111111	ппппп	111111111		1111111111	
	orf26a	MOLIDYSHSFFSVVI	PPFLALALAV	/ITRRVLLSLG	IGILVGVAFL		HLKDMV
		10	20	30	40	50	60
55							
		70	80	90	99		
	orf26.pep	VGLAWSDXDWSLGK!	PK <u>ILVFXIL</u> I	LGIFTSLLTYS	GSNXX		
		111111 11111	11 111 111		1111		
•	orf26a	VGLAWSDGDWSLGK	PKXLVFLILI	LGIFTSLLTYS			
60		70	80	90	100	110	120

	orf26.pep					
5	orf26a	LVFVTFIDDYFHSLA		KVSRAKLAYILD 150 16		
10	orf26.pep					
10	orf26a	TLAGLLVTYKITEYT		YYALFALIMVFV 210 22		
15	orf26.pep				100	110 TSLV
•	orf26a	AHDETAVSDGSWGRV		STVSAMIYTGAQ 270 28		 ENTDVNTSLV 300
20						
20	orf26.pep	120 130 FGGTCGVFAVVLCTI	GTIKTADYPKAV	150 WQGAKSMFGAIA	160 ILILAWLISTV	170 VGEMHTGDYL
	orf26a	FGGTCGVLAVVLCTI	<u>L</u> GTIKIADYPKAV	WOGAKSM <u>FGAIA</u> 330 34	ILILAWLISTV	VGEMHTGDYL
25						360
•	orf26.pep	180 190 STLVAGNIHPGFLPV	/ILFLLASVMAFA	210 TGTSWGTFGIML	220 PIAAAMAVKVE	230 P <u>ALIIPCMSA</u>
20	orf26a	STLVAGNIHPGFLX	/ILFLLASVMAFA	rgtsw <u>gtfgiml</u>	PIAAAMAVKVD	:         PSLIIPCMSA
30		370	380	390 40	0 410	420
	orf26.pep	240 250 VMAGAVCGDHCSPIS		270 NHIDHVTSOLPY	280 ALTVAAAAASG	290 YLALGI.TKSA
35	orf26a	VMAGAVCGDHCSPIS		,,,,,,,,,,,,,	11111111	11111111
		430		450 46		480
	orf26.pep	300 310				
40	orf26a	:           LLGFGXTGIVLAVL	ПП			
	011200	490	500			
	The complete leng	th ORF26a nucleo	tide sequence	<seq 693<="" id="" th=""><th>&gt; is:</th><th></th></seq>	> is:	
45	1 A7 51 T7	GCAGCTGA TCGACTA	ATTC ACATTCAT	TT TTCTCGGTT	G TGCCACCCT	ŗ
1.5	101 G0	CATCGGTAT TCTGGT(	GGC GTTGCCTT	TT TGGTCGGCG	G CAACCCCGT	2
	151 GP 201 CC	ACGGTCTGA CACACCT GCGATTGG TCGCTGG	GAA AGACATGG'	TC GTCGGCTTG	G CTTGGTCAG	A.
50	251 TO	GGTATTTT TACTTCO	CTG CTGACCTA	CT CCGGCAGCA	A TCAGGCGTT'	r ·
50	301 GG 351 GA	CCGACTGGG CAAAACG ACCGCCTGC CTCGTG1	GCA CATTAAAA TTCG TAACCTTT	AC CGGCGCGGC	G CGAAAATGC	r ·
	401 TO	CGCCGTCGG TGCGNTT	GCC CGCCCCGT	PA CCGACAAGT	T TAAAGTTTC	2
	451 CC 501 GC	ECGCCAAAC TCGCCTA CTGATGCCC GTTTCAA	ACAT CCTCGACT	CC ACTGCCGCG	C CTATGTGCG	r
55	551 G <i>I</i>	ACTGCTCGT TACCTAC	CAAA ATCACCGA	AT ACACGCCGA	T GGGGACGTT'	r
	601 GT 651 GT	CGCCATGA GCCTGAT TCGTCGTC GCATGGT	GAA CTATTACG	CA CTGTTTGCC	C TGATTATGG	r
	701 AZ	ACAAGCCGC GTTGAAC	GAA GCCCACGA	IG AAACTGCCG	T TTCAGACGG	2
60	751 AC	SCTGGGGCA GGGTTT#	ACGC ATTGATTA	TT CCCGTTTTG	G CCTTAATCG	2
00	851 TO	CAACGGTT TCCGCCA CAGCATTTT GGGTGCA	ATTT GAAAATAC	GG TGCACAGGC. GG ACGTGAACA	A AGCGAAACC' C TTCGCTGGT	r A
	901 T	CGGCGGCA CTTGCGG	CGT GCTTGCCG	IC GTCCTCTGC	A CGCTCGGCA	3
	951 GA 1001 TO	ATTAAAATC GCCGATT STTCGGCGC AATCGCC	TATU CCAAAGCC CATT TTAATCCT	FT TTGGCAGGG FG CCTGGCTCA	T GCGAAATCC T CAGTACGGT	A r
65	1051 G1	rcgccaaa tgcacac	CAGG CGACTACC	IC TCCACGCTG	G TTGCGGGCA	A
	1101 CA 1151 TO	ATCCATCCC GGCTTCC GGCGTTTGC CACAGGC	TGN CCGTCATC	CT TTTCCTGCT	C GCCAGCGTG	A
	1201 A	TTGCCGCCG CCATGG(	GGT CAAAGTCG	AT CCCTCACTG	A TTATCCCGT	3
70	1251 TA	ATGTCCGCC GTGATGO TTCCGACAC GACCATO	SCGG GGGCGGTA	IG CGGCGACCA	C TGCTCGCCC	A
. •		Journa Groom	ICGICCAC	ce ecececect	G CHACCACAT	-

PCT/IB98/01665

10

15

		-397-
	1401 CGCAT 1451 TTGGC 1501 AAAAA	ACGTTA CNTCGCAACT GCCTTACGCC TTAACCGTTG CCGCCGCCGC CCGGGN TACCTCGCAT TGGGTCTGAC AAAATCCGCG CTGTTGGGTT CANGAC AGGCATTGTA TTGGCGGTGC TGATTTTCT GTTGAAAGAT AACGCG CCAACGCCTG A
5	This encodes a protein	having amino acid sequence <seq 694="" id="">:</seq>
10 15	51 DGLTI 101 ADWAI 151 RAKL 201 VAMS: 251 SWGR' 301 FGGT' 351 VGEM 401 TAAA 451 DHVT 501 KKRA	
	ORF26a and ORF26-	1 show 97.8% identity in 506 aa overlap:
20	orf26a.pep orf26-1	10 20 30 40 50 60  MQLIDYSHSFFSVVPPFLALALAVITRRVLLSLGIGILVGVAFLVGGNPVDGLTHLKDMV
25	orf26a.pep	70 80 90 100 110 120  VGLAWSDGDWSLGKPKXLVFLILLGIFTSLLTYSGSNQAFADWAKRHIKNRRGAKMLTAC
30	orf26a.pep orf26-1	130 140 150 160 170 180 LVFVTFIDDYFHSLAVGAXARPVTDKFKVSRAKLAYILDSTAAPMCVLMPVSSWGASIIA
35	orf26a.pep	190 200 210 220 230 240
40	orf26-1	I AGBLIVITY   I   I   I   I   I   I   I   I   I
	orf26a.pep	250 260 270 280 290 300 AHDETAVSDGSWGRVYALIIPVLALIASTVSAMIYTGAQASETFSILGAFENTDVNTSLV         ::
45	orf26-1	250 260 270 280 290 300
50	orf26a.pep orf26-1	310 320 330 340 350 360 FGGTCGVLAVVLCTLGTIKIADYPKAVWQGAKSMFGAIAILILAWLISTVVGEMHTGDYL
55	orf26a.pep orf26-1	370 380 390 400 410 420 STLVAGNIHPGFLXVILFLLASVMAFATGTSWGTFGIMLPIAAAMAVKVDPSLIIPCMSA
60	orf26a.pep orf26-1	430 440 450 460 470 480 VMAGAVCGDHCSPISDTTILSSTGARCNHIDHVTSQLPYALTVAAAAASGYLALGLTKSA
		430 440 450 460 470 480

65 490 500 LLGFGXTGIVLAVLIFLLKDKKRANAX orf26a.pep

orf26-1

# 5 Homology with a predicted ORF from N. gonorrhoeae

ORF26 shows 94.8% and 99% identity in 97 and 206 as overlap at the N-terminus and C-terminus, respectively, with a predicted ORF (ORF26ng) from N. gonorrhoeae:

	Teshecutory,		
	orf26.pep	MQLIDYSHSFFSVVPPFLALALAVITRRVLLSLGIGILXXVAFLVGGNPVDGLTHLKDMV	60 .
			60
10	orf26ng		97
	orf26.pep	VGLAWSDXDWSLGKPKILVFXILLGIFTSLLTYSGSN	120
	orf26ng	:	120
15		. //	•
		TSLVFGGTCGVFAVVLCTLGTIKTADYPKA	326
•	orf26.pep	111111111111111111111111111111111111111	326
20	orf26ng	ASTVSAMIYTGAQASETFSILGAFENTDVNTSLVFGGTCGVLAVVLCTFGTIKTADYPKA	
	orf26.pep	VWQGAKSMFGAIAILILAWLISTVVGEMHTGDYLSTLVAGNIHPGFLPVILFLLASVMAF	386 386
	orf26ng	VWQGAKSMFGAIAILILAWLISTVVGEMHTGDYLSTLVAGNIAFGF BF VIBLAMIO VILL	
25	orf26.pep	ATGTSWGTFGIMLPIAAAMAVKVEPALIIPCMSAVMAGAVCGDHCSPISDTTILSSTGAR	446 446
	orf26ng	ATGTSWGTFGIMLPIAAAMAVKVEPALIIPCMSAVMAGAVCGBMCSIISSII	1 1
30	orf26.pep	CNHIDHVTSQLPYALTVAAAAASGYLALGLTKSALLGFGTTGIVLAVLIFLLKDKK	502 506
	orf26ng	CNHIDHVTSQLPYALTVAAAAASGYLALGLTKSALLGFGTTGIVLAVLIFLLKDKKRADV	300

The complete length ORF26ng nucleotide sequence <SEQ ID 695> is:

	1	ATGCAGCTGA '	TTGACTATTC	ACATTCATTT	TTCTCGGTTG	TGCCACCCTT
0.5	51		へへみとかかだとして	TOATTALLE	CCGCGTWCTG	0101011111
35	101		BERCOROCCO	<b>Շ</b> ատշ ( Մահանուհայ և	TUBLICUGU	CHACCCCGIC
	151		~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~	አሮአሮአጥርርጥር	GTCGGCTTGG	CIIGGGCAGA
,	201		mcccmcccca.	<b>DACCABABAT</b>	CTTGGTTLIC	CIGNINCILI
	251			CTCACCTACT	CCGGGGGGGAA	ICEGCCIII
40	301		CANARACCCCA	$C\Delta TT\Delta \Delta \Delta \Delta \Delta \Delta C$	CGGTGCGGCG	CGWWWIGCI
40	351			ጥ እ እ ር ርጥጥ ነገ A ነገ	CGACGACIAI	TICCHCROCC
	401		MACCA MINCOCO	CCCCCCGTTA	LICHALANGII	IMMOITICO
	451		$\mathbf{m} \mathbf{c} \mathbf{c} \mathbf{c} \mathbf{c} \mathbf{c} \mathbf{m} \mathbf{x} \mathbf{c} \mathbf{x} \mathbf{m}$	CCTCGACTCC	ACTGCCTCGC	CCVIGIOCOI
	501			CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	GATTATUGUU	MCGCIIGCCG
AE	551		$m_{\lambda} \wedge \wedge m_{\lambda} \wedge \gamma_{\lambda} \lambda_{\lambda}$	- አጥጥአሮርGAA'I'	ACACGCCGAI	GGGGGCGIII
45	601		CCCTCTTCTTCTT	CTATTACGCG	CTGTTTGCCC	IGNITATEGI
•	651	- mmacmacma	こころ 中にこの中でで	CCTTCGACAT	CGGCTCGACG	gcacarrica
	701		<b>~四四~カカ~~カカ</b>	- acccaddacd	aaaccuccuc	CICRORCACI
	751	2 CC 2 2 2 CC CC	CMCMMMDCCC	ATTGATTATT	CCCGTTTIGG	CCITALICOC
50	801		macccconfica	- ጥርጥ <b>አ</b> ሮአርርርር	CGCGCAGGCA	MGCGWWYCCI
50	851		へへへへへへへ入中中中	CAAAATACCG	ACGIAAACAC	TICGCIGGIA
	901		COMPACACACA	COUNTIES	GICCICIGCA	CGIICGGCWC
	951		- へへへへ カ 中 中 カ 中 (	' CCAAAGCCCG'I	'GIGGCAGGGI	GCGWWICCW
	1001		*************************************	י ጥጥልልጥሮርጥር0	CCTGGCTCAT	CAGIACGGII
E E	1051		かいこう こうじょう	: CCACTACCTO	: TCCACGCTGC	TIGCGGGCAA
55	1101	02 m 0 0 2 m 0 0 0	-CCCTTCCTCC	CCCTCATCC	CTTCCTGCTC	, GCCAGCGIGA
	1151		CACACCCAC1	<u> </u>	A CGTTCGGCA	TAIGCIGCCG
	1201			ר כאאאכירכא	A CCCGCGCTGA	ITMCCCCGIG
	1251		・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・	: CCCCCGTAT(	3 CGGCGACCA	, IGIICGCCCA
<b>C</b> D	1301		• උአረረአጥርርጥ(	TOGTOCACO	3 6666666611	3 CHACCHORIC
60	1351			r cccmmatgc	CTGACGGIIV	3 CCGCCGCCGC
	1401		* MXCCTCCCX	т тасатта	C AAAATCCGC	e Cigiiagaci
	1451		CGGTATTGT	A TTGGCGGTG	C TGATTTTTC	T GTTGAAAGAT
	1501		CCGACGTTT	G A		•
	1001				<\$EO ID 69	۲.
	_			id accused	くくらい けいかり	0/:

This encodes a protein having amino acid sequence <SEQ ID 696>:

STLVAGNIHPGFLPVILFLLASVMAFATGTSWGTFGIMLPIAAAMAVKVEPALIIPCMSA 

STLVAGNIHPGFLPVILFLLASVMAFATGTSWGTFGIMLPIAAAMAVKVEPALIIPCMSA

VMAGAVCGDHCSPISDTTILSSTGARCNHIDHVTSQLPYALTVAAAAASGYLALGLTKSA

VMAGAVCGDHCSPISDTTILSSTGARCNHIDHVTSQLPYALTVAAAAASGYLALGLTKSA

orf26-1.pep

orf26-1.pep

orf26-1.pep

orf26ng

orf26ng

orf26ng

			LCI/IRA9/01002
	WO 99/24578	200	
		-399-	
	1 40	LIDYSHSF FSVVPPFLAL ALAVITRRVL LSLGIGILVG VAFLV	GGNPV
		THE PERSON AND A PERSON DELIGIOUS OF CANDRELL OF THE PERSON DESIGNATIONS OF THE PERSON DELICITIES OF THE PERSON DELICITIE	0110211
			MERCO
			LITUIT
-		COLONY TENTIMIEN AWESFOLGSM ARTEUAALNE AUDEL	MAGUA .
5			/ 1.1 = U = 1.
			1DTOTA
		omitte Cut DUTI VIII ASVMATATUL SWULL	GILILI
			71/01/117
10	451 DH	VTSQLPYA LTVAAAAASG YLALGLTKSA LLGFGTTGIV LAVL	LEBERD
10	501 KK	RADV*	
		no C 1 . 1 00 40/ identity in 505 as overlan	
	ORF26ng and ORF	26-1 show 98.4% identity in 505 aa overlap:	
	_	40	50 60
		10 20 30 40 MQLIDYSHSFFSVVPPFLALALAVITRRVLLSLGIGILVGVAF	
	orf26-1.pep		
15		MQLIDYSHSFFSVVPPFLALALAVITRRVLLSLGIGILVGVAF	LVGGNPVDGLTHLKDMV
	orf26ng	10 20 30 40	50 60
		10 20 30	
		70 80 90 100	110 120
	50.5 1	THE PROPERTY OF THE THE STATE OF THE STATE O	AKRHIKNRRGAKMLTAC
20	orf26-1.pep		
	52629	VGLAWADGDWSLGKPKILVFLILLGIFTSLLTYSGSNQAFADV	MVKUTKUKCOMAM
	orf26ng	70 80 90 100	110 120
		1.50	170 180
25		130 140 150 160	
23	orf26-1.pep	LVFVTFIDDYFHSLAVGAIARPVTDKFKVSRTKLAYILDSTA	· · · · · · · · · · · · · · · · · · ·
	• • •	LVFVTFIDDYFHSLAVGATARFVIDARKVSKILLIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	SPMCVIMPVSSWGASIIA
	orf26ng		170 180
		130 140 150 160	-
30		190 200 210 220	230 240
		190 200 MNYVAI.FAI.TMVFVVAW	FSFDIGSMARFEQAALNE
	orf26-1.per		1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
	52655	TI.AGI.I.VTYKITEYTPMGTFVAMSLMNYYALFALIMVEVVAM	I DI DI COLLINIZIONI
25	orf26ng	190 200 210 220	230 240
35			290 300
		250 260 270 280	
	orf26-1.pe	AHDETAVSDATKGRVYALIIPVLALIASTVSAMIYTGAQASE	TESTEGAPENIDANISMA
	01120		
40	orf26ng	AQDETAASDATKGRVYALIIPVLALIASTVSAMIYTGAQASE	290 300
70		250 260 270 280	230
		310 320 330 340	350 360
		310 320 STUDY OF CONTROL OF CONTR	LAWLISTVVGEMHTGDYL
	orf26-1.pe		
45	ه ه د	FGGTCGVLAVVLCTFGTIKTADYPKAVWQGAKSMFGAIAIL	THUMBIOISSON
	orf26ng	310 320 330 340	350 360

In addition, ORF26 ng shows significant homology to a hypothetical H.influenzae protein:

LLGFGTTGIVLAVLIFLLKDKKRANAX

LLGFGTTGIVLAVLIFLLKDKKRADVX

```
sp|P44263|YF86_HAEIN
                                  HYPOTHETICAL
                                                 PROTEIN
                                                           HI1586
                                                                     >gi|1074850|pir||C64037
          hypothetical
          protein HI1586 - Haemophilus influenzae (strain Rd KW20) >gi|1574427 (U32832) H.
          influenzae predicted coding region HI1586 [Haemophilus influenzae] Length = 519
 5
            Score = 538 \text{ bits (1370)}, Expect = e-152
            Identities = 280/507 (55%), Positives = 346/507 (68%), Gaps = 7/507 (1%)
                      {\tt MQLIDYSHSFFSVVPPFLALALAVITRRXXXXXXXXXXXXXXXAFLVGGNPVDGLTHLKDMV~60}
                      M+LID+S S +S+VP LA+ LA+ TRR
10
                      MELIDFSSSVWSIVPALLAIILAIATRRVLVSLSAGIIIGSLMLSDWQIGSAFNYLVKNV 73
          Sbjct: 14
           Query: 61 VGLAWADGDWSLGKPKILVFLILLGIFTSLLTYSGSNQAFADWAKRHIKNRCGAKMLTAC 120
                      V L +ADG+ +
                                      I++FL+LLG+ T+LLT SGSN+AFA+WA+ IK R GAK+L A
           Sbjct: 74 VSLVYADGEIN-SNMNIVLFLLLLGVLTALLTVSGSNRAFAEWAQSRIKGRRGAKLLAAS 132
15
           Query: 121 LVFVTFIDDYFHSLAVGAIARPVTDKFKVSRAKLAYILDSTASPMCVLMPVSSWGASIIA 180
                      LVFVTFIDDYFHSLAVGAIARPVTD+FKVSRAKLAYILDSTA+PMCV+MPVSSWGA II
           Sbjct: 133 LVFVTFIDDYFHSLAVGAIARPVTDRFKVSRAKLAYILDSTAAPMCVMMPVSSWGAYIIT 192
20
           Query: 181 TLAGLLVTYKITEYTPMGTFVAMSLMNYYALFALIMVFVVAWFSFDIGSMARFEQAALNE 240
                       + GLL TY ITEYTP+G FVAMS MN+YA+F++IMVF VA+FSFDI SM R E+ AL
           Sbjct: 193 LIGGLLATYSITEYTPIGAFVAMSSMNFYAIFSIIMVFFVAYFSFDIASMVRHEKLALKN 252
           Query: 241 AQDETAASDATKGRVYALIIPVLALIASTVSAMIYTGAQA----SETFSILGAFENTDVN 296
25
                                TKG+V LI+P+L LI +TVS MIYTGA+A
                       +D+
                                                                    + FS+LG FENT V
           Sbjct: 253 TEDQLEEETGTKGQVRNLILPILVLIIATVSMMIYTGAEALAADGKVFSVLGTFENTVVG 312
           Query: 297 TSLVFGGTCGVL--AVVLCTFGTIKTADYPKAVWQGAKSMFGXXXXXXXXXXXXXXXTVVGEM 354
                      TSLV GG C ++ +++
                                            + +Y ++
                                                         G KSM G
30
           Sbjct: 313 TSLVVGGFCSIIISTLLIILDRQVSVPEYVRSWIVGIKSMSGAIAILFFAWTINKIVGDM 372
           Query: 355 HTGDYLSTLVAGNIHPGFLPVILFLLASVMAFATGTSWGTFGIMLPIAAAMAVKVEPALI 414
                       TG YLS+LV+GNI
                                      FLPVILF+L + MAF+TGTSWGTFGIMLPIAAAMA
                                                                               P L+
           Sbjct: 373 QTGKYLSSLVSGNIPMQFLPVILFVLGAAMAFSTGTSWGTFGIMLPIAAAMAANAAPELL 432
35
           Query: 415 IPCMSAVMAGAVCGDHCSPISDTTILSSTGARCNHIDHVTSQXXXXXXXXXXXXXXXXXX 474
                      +PC+SAVMAGAVCGDHCSP+SDTTILSSTGA+CNHIDHVT+Q
           Sbjct: 433 LPCLSAVMAGAVCGDHCSPVSDTTILSSTGAKCNHIDHVTTQLPYAATVATATSIGYIVV 492
40
           Query: 475 XXXKSALLGFGTTGIVLAVLIFLLKDK 501
                          S L GF T + L V+IF +K +
           Sbjct: 493 GFTYSGLAGFAATAVSLIVIIFAVKKR 519
```

Based on this analysis, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 83

45

The following partial DNA sequence was identified in *N. meningitidis* <SEQ ID 697>:

```
1 ..AAGCAATGGT ATGCCGACGN .AGTATCAAG ACGGAAATGG TTATGGTCAA
51 CGATGAGCCT GCCAAAATTC TGACTTGGGA TGAAAGCGGC CGATTACTCT
50 101 CGGAACTGTC TATCCGCCAC CATCAACGCA ACGGGGTGGT TTTGGAGTGG
151 TATGAAGATG GTTCTAAAAA GAGCGAAGT. GTTTATCAGG ATGACAAGTT
201 GGTCAGGAAA ACCCAGTGGG ATAAGGATGG TTATTTAATC GAACCCTGA
```

This corresponds to the amino acid sequence <SEQ ID 698; ORF27>:

1 ..KQWYADXSIK TEMVMVNDEP AKILTWDESG RLLSELSIRH HQRNGVVLEW
55 51 YEDGSKKSEX VYQDDKLVRK TQWDKDGYLI EP\*

Further work revealed the complete nucleotide sequence <SEO ID 699>:

```
1 ATGAAAAAT TATCTCGGAT TGTATTTCA ACTGTCCTGT TGGGTTTTTC
51 GGCCGCTTG CCGGCGCAGA CCTATTCTGT TTATTTTAAT CAGAACGGAA
101 AGCTGACGGC GACGATGTCT TCTGCCGCTT ATATCAGGCA ATATAGTGTG
60 151 GTGGCGGGTA TTGCGCACGC GCAGGATTTT TATTATCCGT CGATGAAGAA
```

	-401-
5	201 ATATTCTGAA CCTTATATCG TTGCTTCAAC GCAAATCAAA TCTTTTGTGC 251 CTACCCTGCA AAACGGTATG TTGATTTTGT GGCATTTTAA TGGTCAGAAA 301 AAAATGGCGG GGGGCTTCAG CAAGGGTAAG CCGGACGGGG AGTGGGTCAA 351 CTGGTATCCG AACGGTAAAA AATCTGCCGT TATGCCTTAT AAAAATGGCT 401 TGAGTGAGGG TACGGGATAC CGCTATTACC GTAACGGCGG CAAGGAAAGC 451 GAAATCCAGT TTAAGCAAAA TAAGGCAAAC GGCGTATGGA AGCAATGGTA 501 TGCCGACGGC AGTATCAAGA CGGAAATGGT TATGGTCAAC GATGAGCCTG 551 CCAAAATTCT GACTTGGGAT GAAAGCGGCC GATTACTCTC GGAACTGTCT 601 ATCCGCCACC ATCAACGCAA CGGGGTGGTT TTGGAGTGGT ATGAAGATGG 651 TTCTAAAAAG AGCGAAGCTG TTTATCAGGA TGACAAGTTG GTCAGGAAAA 701 CCCAGTGGGA TAAGGATGGT TATTTAATCG AACCCTGA
	This corresponds to the amino acid sequence <seq 700;="" id="" orf27-1="">:</seq>
15	1 MKKLSRIVFS TVLLGFSAAL PAQTYSVYFN QNGKLTATMS SAAYIRQYSV 51 VAGIAHAQDF YYPSMKKYSE PYIVASTQIK SFVPTLQNGM LILWHFNGQK 101 KMAGGFSKGK PDGEWVNWYP NGKKSAVMPY KNGLSEGTGY RYYRNGGKES 151 EIQFKQNKAN GVWKQWYADG SIKTEMVMVN DEPAKILTWD ESGRLLSELS 201 IRHHQRNGVV LEWYEDGSKK SEAVYQDDKL VRKTQWDKDG YLIEP*
	Computer analysis of this amino acid sequence gave the following results:
	Homology with a predicted ORF from N. meningitidis (strain A)
20	ORF27 shows 91.5% identity over a 82aa overlap with an ORF (ORF27a) from strain A of N.
	meningitidis:
	10 20 30
	orf27.pep KQWYADXSIKTEMVMVNDEPAKILTWDESG
25	orf27a LSEGTGXRYYRNGGKESEIQFKQNKANGVWKQWYADGNIKTEMVMVNDEPAKILTWDESG 140 150 160 170 180 190
	40 50 60 70 80
30	orf27.pep RLLSELSIRHHQRNGVVLEWYEDGSKKSEXVYQDDKLVRKTQWDKDGYLIEPX
50	orf27a RLLSELSIHHHXRNGVVLEWYEDGSKKXEAVYQDDKLVRKTQWDXDGYLIEPX 200 210 220 230 240
	The complete length ORF27a nucleotide sequence <seq 701="" id=""> is:</seq>
35	51 GGCCGCTTTG CCGGCGCAGA NCTATTCTGT TTATTTTAAT CAGAACGGGA
	151 GCGGAGGGTA TTGCGCACGC GCAGGANTTT TANTATCCGT CGATGAAGAA
	201 ATATTCCGAA CCTTATATCG TTGCTTCAAC GCAAATCAAA TCTTTTGTGC 251 CTACCCTGCA AAACGGTATG TTGATTTTGT GGCATTTTAA NGGTCAGAAA
40	301 AAAATGGCNG GGGGCTTCAG CAAGGGTAAG CCGGACGGGG AGTGGGTCAA 351 CTGGTATCCG AACGGTAAAA AATCTGCCGT TATGCCTTAT AAAAATGGTT
	401 TGAGTGAAGG TACGGGGTNN CGCTATTACC GTAACGGCGG CAAGGAAAGC 451 GAAATCCAGT TTAAACAGAA TAAGGCAAAC GGCGTATGGA AGCAATGGTA
45	501 TGCCGACGGC AATATCAAAA CGGAAATGGT TATGGTCAAT GATGAGCCTG 551 CCAAAATTCT GACATGGGAT GAAAGCGGTC GATTACTCTC GGAACTGTCT
	601 ATCCATCATC ATNAACGTAA TGGAGTAGTC TTAGAGTGGT ATGAAGATGG 651 TTCTAAAAAG ANTGAAGCTG TTTATCAGGA TGATAAGTTG GTCAGGAAAA
	701 CCCAGTGGGA TAANGATGGT TATTTAATCG AACCCTGA
	This encodes a protein having amino acid sequence <seq 702="" id="">:</seq>
50	1 MKKLSRIVFS TVLLGFSAAL PAQXYSVYFN QNGKLTATXS SAAYIRQYSV 51 AEGIAHAQXF XYPSMKKYSE PYIVASTQIK SFVPTLQNGM LILWHFXGQK
	101 KMAGGFSKGK PDGEWVNWYP NGKKSAVMPY KNGLSEGTGX RYYRNGGKES
	151 EIQFKQNKAN GVWKQWYADG NIKTEMVMVN DEPAKILTWD ESGKLLSELS 201 IHHHXRNGVV LEWYEDGSKK XEAVYQDDKL VRKTQWDXDG YLIEP*
55	ORF27a and ORF27-1 show 94.7% identity in 245 aa overlap:
	10 20 30 40 50 60

				•			
	orf27-1		:     SAALPAQTYS   20	 VYFNQNGKLTA 30	II IIIIIII ATMSSAAYIR 40	:     QYSVVAGIA 50	HAQDF 60
5		70	80	90	100	110	120
	orf27a.pep	XYPSMKKYSEPYIVAS	STQIKSFVPTL(	QNGMLILWHF?	KGQKKMAGGF	SKGKPDGEW	VNWYP
	orf27-1	YYPSMKKYSEPYIVAS	STQIKSFVPTL	ONGMLILWHFT	NGOKKMAGGE 100	SKGKPDGEW	VNWYP
10							120
	orf27a.pep	130 NGKKSAVMPYKNGLSE	140 EGTGXRYYRNG	150 GKESEIQFKQ1	160 NKANGVWKQW	170 YADGNIKTE	180 MVMVN
	orf27-1			11111111111		1111:1111	HIII
15	·	130	140	150	160	170	180
		190	200	210	220	230	240
	orf27a.pep	DEPAKILTWDESGRLI	SELSIHHHXRI	NGVVLEWYEDO	SSKKXEAVY <u>C</u> 	DDKLVRKTQ	WDXDG
20	orf27-1	DEPAKILTWDESGRLI 190	SELSIRHHQRI 200	NGVVLEWYEDO	SSKKSEAVYÇ	DDKLVRKTQ	WDKDG
		190	200	210	220	230	240
	orf27a.pep	YLIEPX					
25	orf27-1	 YLIEPX				. ,	
		district ODE C	T 1				
	Homology with a pre			•		•	i .
	ORF27 shows 96.3%	6 identity over 82	aa overlap	with a pr	redicted O	RF (ORF	27ng) from
30	N.gonorrhoeae:	•		,			
	orf27.pep				KTEMVMVND		
	orf27ng I	SEGTGYRYYRNGGKESE	IQFKQNKANG	 :VWKQWYADGS		 EPAKILTWD	 ESG 193
35		LLSELSIRHHQRNGVVI					
	J		1111111111	F 111111111	1111111111	1111	
				*		LIEP 245	
	The complete length	OKF2/lig nucleon	ie sequence	~2EQ ID /	03> is:		
40	1 ATGA 51 GGCC	AGAAAT TATCTCGGAT	TGTATTTTC	A ATCGTACTO	T TGGGTTT	TTC	
40	101 AACT	GCTTTG CCGGCGCAGA GACGGC GACGATGTCT	TCTGCCGCT	T ATATCAGG	CA ATATAGT	GTG	
	151 GCGG 201 ATAT	CGGGTA TCGCACACGC	CAGGATTT: TTGCTTCAA	T TATTATCC( C GCAAATCA	ST CGATGAA AA TCTTTTG	GAA TGC	
45	251 CTAC	CCTGCA AAACGGTATG TGGCGG GGGGCTTCAG	TTGATTTTG'	T GGCATTTT	AA TGGTCAG	AAA .	•
	351 CTGG	TATCCG AACGGTAAAA	AATCTGCGG	TATGCCTTA	TAAAAAT	GCT	
	451 GAAA	TGAGGG TACGGGATAC TCCAGT TTAAGCAAA	TAAGGCGAA	C GGCGTATGC	A AGCAATG	GTA	
50	501 TGCC	GATGGA AGTATCAAGA AATTCT GACTTGGGAT	CGGAAATGG	TATGGTCA	AC GATGAGO	CTG	
	601 ATCC	GCCACC ATAAACGCAA	CGGGGTGGT'	T TTGGAGTG	T ATGAAGA	TGG	
	651 TTCT 701 CCCA	'AAAAAG AGCGAGGCTG ATGGGA TAAGGATGG1	TTTATCAGG	A TGACAAGTI G AACCCTGA	rg gtcagga	AAA	
	This encodes a protei	n having amino aci	d sequence	<seo 70<="" id="" th=""><th>)4&gt;:</th><th></th><th></th></seo>	)4>:		
55				_			
55	1 MKKI 51 AAGI	SRIVFS IVLLGFSAAI AHAQDF YYPSMKKYSE	PAQTYSVYFI PYIVASTOII	N QNGKLTATI K SFVPTLONG	AS SAAYIRQ M LILWHFN	YSV GOK	
	101 KMAG	GFSKGK PDGEWVNWYE KQNKAN GVWKQWYADO	P NGKKSAVMP	Y KNGLSEGTO	Y RYYRNGO	KES	
	201 IRHH	KRNGVV LEWYEDGSKK	SEAVYQDDK	L VRKTQWDKI	OG YLIEP*	· ELS	
60	ORF27ng and ORF27	7-1 show 98.8% ide	entity in 245	aa overlap:			
		10	20	30	40	50	60
	orf27-1.pep	MKKLSRIVFSTVLLGE	SAALPAQTYS'	VYFNQNGKLT!	ATMSSAAYIR	QYSVVAGIA	HAQDF

	orf27ng		  DF  60
5	orf27-1.pep	YYPSMKKYSEPYIVASTOIKSFVPTLONGMLILWHFNGOKKMAGGFSKGKPDGEWVNW	i I I
	orf27ng	YYPSMKKYSEPYIVASTQIKSFVPTLQNGMLILWHFNGQKKMAGGFSKGKPDGEWVNW 70 80 90 100 110 1	120
10	orf27-1.pep	NGKKSAVMPYKNGLSEGTGYRYYRNGGKESEIQFKQNKANGVWKQWYADGSIKTEMVN	180 MVN
15	orf27ng	NCKKSAVMPYKNGLSEGTGYRYYRNGGKESEIOFKONKANGVWKQWYADGSIKTEMV	
	orf27-1.pep	DEPAKILTWDESGRLLSELSIRHHQRNGVVLEWYEDGSKKSEAVYQDDKLVRKTQWD	111
20	orf27ng	DEPAKILTWDESGRLLSELSIRHHKRNGVVLEWYEDGSKKSEAVYQDDKLVRKTQWD 190 200 210 220 230	KDG 240
25	orf27-1.pep	YLIEPX 	

Based on this analysis, including the putative leader sequence in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF27-1 (24.5kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 17A shows the results of affinity purification of the GST-fusion protein, and Figure 17B shows the results of expression of the His-fusion in *E.coli*. Purified GST-fusion protein was used to immunise mice, whose sera were used for ELISA, which gave a positive result, confirming that ORF27-1 is a surface-exposed protein and a useful immunogen.

#### Example 84

30

35

The following partial DNA sequence was identified in N. meningitidis <SEQ ID 705>:

	1	ATGAAATTTA	CCAAGCACCC	CGTCTGGGCA	ATGGCGTTCC	GCCCATTTTA
40	51	TTCGCTGGCG	GCTCTGTACG	GCGCATTGTC	CGTATTGCTG	TGGGGTTTCG
70	101	GCTACACGGG	AACGCACKAG	CTGTCCGGTT	TCTATTGGCA	CGCGCATGAg
	151	ATGATTTGGG	GTTATGCCGG	ACTGGTCGTC	ATCGCCTTCC	TGCTGACCGC
	201	CGTCGCCACT	TGGACGGGGC	AGCCGCCCAC	GCGGGGCGGC	GTaTCTGGTC
	251	GGCTTGACTA	TCTTTTGGCT	GGCTGCGCGG	ATTGCCGCCT	TTATCCCGGG
45	301	TTGGGGTGCG	TCGGCAAGCG	GCATACTCGG	TACGCTGTTT	TTCTGGTACG
7.7	351	CCCCCCTCTC	CATGGCTTTG	CCCGTTATCC	GTTCGCAGAA	TCAACGCAAC
	401	TATGTTGCCG	TGTTCGCGCT	GTTCGTCTTG	GGCGGCACGC	ATGCGGCGTT
	451	CCACGTCCAG	CTGCACAACG	GCAACCTAGG	CGGACTCTTG	AGCGGATTGC
	501	AGTCGGGCTT				

50 This corresponds to the amino acid sequence <SEQ ID 706; ORF47>:

<sup>1</sup> MKFTKHPVWA MAFRPFYSLA ALYGALSVLL WGFGYTGTHX LSGFYWHAHE 51 MIWGYAGLVV IAFLLTAVAT WTGQPPTRGG VLVGLTIFWL AARIAAFIPG

<sup>101</sup> WGASASGILG TLFFWYGAVC MALPVIRSON ORNYVAVFAL FVLGGTHAAF

151 HVQLHNGNLG GLLSGLQSGL VM

# Further work revealed the complete nucleotide sequence <SEQ ID 707>:

		1	ATGAAATTTA	CCAAGCACCC	CGTCTGGGCA	ATGGCGTTCC	GCCCATTTTA
_		51	TTCGCTGGCG	GCTCTGTACG	GCGCATTGTC	CGTATTGCTG	TGGGGTTTCG
5		101	GCTACACGGG	AACGCACGAG	CTGTCCGGTT	TCTATTGGCA	CGCGCATGAG
		151	ATGATTTGGG		ACTGGTCGTC		
		201	CGTCGCCACT	TGGACGGGGC	AGCCGCCCAC	GCGGGGCGGC	GTTCTGGTCG
	. •	251	GCTTGACTAT		GCTGCGCGGA		TATCCCGGGT
		301			CATACTCGGT		TCTGGTACGG
10		351	CGCGGTGTGC	ATGGCTTTGC	CCGTTATCCG	TTCGCAGAAT	CAACGCAACT
		401	ATGTTGCCGT	GTTCGCGCTG	TTCGTCTTGG	GCGGCACGCA	TGCGGCGTTC
		451	CACGTCCAGC	TGCACAACGG	CAACCTAGGC	GGACTCTTGA	GCGGATTGCA
		501	GTCGGGCTTG	GTGATGGTGT	CGGGTTTTAT	CGGTCTGATT	GGTACGCGGA
- <b>-</b>		551	TTATTTCGTT	TTTTACGTCC	AAACGCTTGA	<b>ATGTGCCGCA</b>	GATTCCCAGT
15		601	CCGAAATGGG	TGGCGCAGGC	TTCGCTGTGG	CTGCCCATGC	TGACTGCCAT
		651	GCTGATGGCG	CACGGTGTGT	TGGCTTGGCT	GTCTGCCGTT	TTTGCCTTTG
		701	CGGCAGGTGT	GATTTTTACC	GTGCAGGTGT	ACCGCTGGTG	GTATAAACCC
		751	GTGTTGAAAG	AGCCGATGCT	GTGGATTCTG	TTTGCCGGCT	ATCTGTTTAC
		801	CGGATTGGGG		TCGGCGCGTC		
20		851	TCAATCTGGG	TGTGCATCTG	ATCGGGGTCG	GCGGTATCGG	CGTGCTGACT
		901			CGCGCTTGGT		
		951	TCCGCCGCCC	AAAGCCGTTC	CCGTTGCGTT	TTGGCTGATG	ATGGCGGCAA
		1001	CCGCCGTCCG	TATGGTTGCC			CTACACGCAC
~ ~		1051			TTTGTTTGCA		
25		1101	GTGGAAGTAT	ATTCCTTGGC	TGATTCGTCC	GCGTTCGGAC	GGCAGGCCCG
		1151	GTTGA	•	•		

# This corresponds to the amino acid sequence <SEQ ID 708; ORF47-1>:

	1	MKFTKHPVWA	MAFRPFYSLA	ALYGALSVLL	WGFGYTGTHE	LSGFYWHAHE
	51	MIWGYAGLVV	IAFLLTAVAT	WTGQPPTRGG	VLVGLTIFWL	AARIAAFIPG
30	101	WGASASGILG	TLFFWYGAVC	MALPVIRSON	QRNYVAVFAL	FVLGGTHAAF
•	151	HVQLHNGNLG	GLLSGLQSGL	VMVSGFIGLI	GTRIISFFTS	KRLNVPQIPS
	201	PKWVAQASLW	LPMLTAMLMA	HGVLAWLSAV	FAFAAGVIFT	VQVYRWWYKP
	251	VLKEPMLWIL	FAGYLFTGLG	LIAVGASYFK	PAFLNLGVHL	IGVGGIGVLT
	301					VFSSGTAYTH
35	351	SIRTSSVLFA				<del>-</del>

Computer analysis of this amino acid sequence predicts a leader peptide and also gave the following results:

# Homology with a predicted ORF from N. meningitidis (strain A)

ORF47 shows 99.4% identity over a 172aa overlap with an ORF (ORF47a) from strain A of N.

## 40 meningitidis:

	orf47.pep	10 MKFTKHPVWAMAFR	20 PFYSLAALY	30 GALSVLLWGFG	40 YTGTHXLSGF	50 YWHAHEMIWG	60 YAGLVV
45	orf47a		 PFYSLAALY			 	 YAGLVV
43		10	20	30	40	50	60
• .		70	80	90	100	110	120
	orf47.pep	IAFLLTAVATWTGO	PPTRGGV <u>LV</u>	<u>GLTIFWLAARI</u>	AAFIPGWGAS	ASGILGTLFF	WYGAVC
50	orf47a	<u>IAFLLTAVATWTGO</u>			AAFI PGWGAS	ASGILGTLFF	WYGAVC
		70	80	90	100	110	120
•		130	140	150	160	170	
55	orf47.pep	MALPVIRSONORNY	VAVFALFVL	GGTHAAFHVQL	HNGNLGGLLS	GLQSGLVM	
55	orf47a	MALPVIRSQNQRNY	VAVFALFVL	GGTHAAFHVQL	HNGNLGGLLS	  GLQSGLVMVS	GFIGLI
		130	140	150	160	170	180
	orf47a	GTRIISFFTSKRLN	VPQIPSPKW	VAQASLWLPML	TAMLMAHGVM	IPWLSAAFAFA	AGVIFT
60		190	200	210	220	230	240

The complete length ORF47a nucleotide sequence <SEQ ID 709> is:

	The complete leng	gm Oid 47a	. Hacicottae .	oquomes o				
	1 7	አመር አ አ አመመሞለ	CCAAGCACCC	ССТТТССССА	ATGGCGTTCC	GCCCGTTTTP	1	
	1 / 51 1	PTCACTGGCG	GCTCTGTACG	GCGCATTGTC	CGTATTGCTG	TGGGGTTTCG	;	
	101 (	GCTACACGGG	AACGCACGAG	CTGTCCGGTT	TCTATTGGCA	CGCGCATGAG	3	
5	151	<b>ልጥርልጥጥጥር</b> ርር	GTTATGCCGG	ACTGGTCGTC	ATCGCCTTCC	TGCTGACCGC	;	
3	201 (	CGTCGCCACT	TGGACGGGGC	AGCCGCCCAC	GCGGGGCGGC	GTTCTGGTC	•	
	251 (	CCTTCACTAT	CTTTTGGCTG	GCTGCGCGGA	TTGCCGCCTT	TATCCCGGGT	ľ	
	301	TGGGGTGCGT	CGGCAAGCGG	CATACTCGGT	ACGCTGTTTT	TCTGGTACGC	<del>,</del>	
	351	CGCGGTGTGC	ATGGCTTTGC	CCGTTATCCG	TTCGCAGAAT	CAACGCAAT	ί ~	
10	401	ATGTTGCCGT	GTTCGCGCTG	TTCGTCTTGG	GCGGTACGCA	CGCGGCGTTC	, ^	
	451	CACGTCCAGC	TGCACAACGG	CAACCTAGGC	GGACTETTGA	CCUACCCCC	<b>1</b>	
	501	GTCGGGCTTG	GTGATGGTGT	CGGGTTTTAT	CGGTCTGATT	CATTCCCAC	3. T	
	551	TTATTTCGTT	TTTTACGTCC TGGCGCAGGC	AAACGGTTGA	CTCCCCATCC	TGACCGCCA'	r P	
	601	CCGAAATGGG	CACGGCGTGA	TTCGCTGTGG	CTCCCCATGC	ですっていることである。	z r	
15	651	GCTGATGGCG	GATTTTTACC	CTCCACCTCT	ACCECTEGTE	GTATAAGCC	ř	
	701 751	CGGCGGGGGGG	AGCCGATGCT	GTGGATTCTG	TTTGCCGGCT	ATCTGTTTA	Ē	
	801	CCCATTCCCC	CTGATTGCGG	TCGGCGCGTC	TTATTTCAAA	CCCGCTTTC	Ċ	
	851	TCAATCTGGG	TGTGCATCTG	ATCGGGGTCG	GCGGTATCGG	CGTGCTGAC	r	
20	901	TTGGGCATGA	TGGCGCGTAC	CGCGCTCGGT	CATACGGGCA	ATCCGATTT.	A	
20	951	TCCGCCGCCC	AAAGCCGTTC	CCGTTGCGTT	TTGGCTGATG	ATGGCGGCA	A	
	1001	CCGCCGTCCG	TATGGTTGCC	GTATTTTCTT	CCGGCACTGC	CTACACGCA	C	
	1051	AGCATACGCA	CCTCTTCGGT	TTTGTTTGCA	CTCGCGCTTT	TGGTGTATG	С	
	1101	GTGGAAGTAT	ATTCCTTGGC	TGATTCGTCC	GCGTTCGGAC	GGCAGGCCC	G ,	
25		GTTGA					:	
				1	CEO ID 710	<b>.</b>		4
	This encodes a p	rotein havin	g amino acio	i sequence <	SEQ ID /II	<i>)</i> :		
	1	MKFTKHPVWA	MAFRPFYSLA	ALYGALSVLL	WGFGYTGTHE	LSGFYWHAH	E S	
	51	MIWGYAGLVV	<u>IAFLLTAV</u> AT	WTGQPPTRGG	VLVGLTIFWI	AARIAAFIP	G F	
	101	WGASASGILG	TLFFWYGAVC GLLSGLQSGL	MALPVIKSON	CWDITEELS	KDINUDOTE	<u>. <del>.</del> </u>	
30		HVQLHNGNLG	LPMLTAMLMA	UCAMBAT 200	FAFAAGVIFT	VOVYRWWYK	P .	
	201	PKWVAQASLW	FAGYLFTGLG	I.TAVGASVEK	PAFINIGVHI	IGVGGIGVI	T	
	301	T CMMAPTALC	HTGNPIYPP	KAVPVAFWIM	MAATAVRMVA	VFSSGTAYT	H	
	351	STDTSSVI.FA	LALLVYAWKY	TPWI.TRPRSE	GRPG*			
35	ORF47a and OR	F47-1 show	99.2% iden	tity in 384 a	a overlap:			
55	010			•	_			
			10	20	30	40	50	60
	orf47a.pe	p MKFTK	HPVWAMAFRPE	YSLAALYGALS	VLLWGFGYTG'	THELSGFYWH?	<b>THEMINGYAG</b>	TVV
	• •	11111	11111111111	111111111111		1111111111	11111111	
	orf47-1	MKFTK	HPVWAMAFRPI			THELSGFYWH!	MEMIWGYAG	ILVV
40			10	20	30	40	50	60
				00	00	100	110	120
			70 TAVATWTGQPI	80 80				
	orf47a.pe	p IAFLI	TAVATWIGQE	PIRGGVLVGLI.	TEMTWRETTUE	IIIIIIIIII	111111111	111
AE			.TAVATWTGQP1		TEWT.AARTAAF	IPGWGASASG:	LLGTLFFWYC	AVC
45	orf47-1	TWE DI	70	80	90	100	110	120
			,0	00	•			
			130	140			170	180
	orf47a.pe	mALP	TRSONORNYV	AVFALFVLGGT	HAAFHVQLHNG	NLGGLLSGLQ	SGLVMVSGFI	[GLI
50		- 1111			11111111111	1111111111	1111111	
50	orf47-1	MALP	/IRSQNQRNYV	AVFALFVLGGT	HAAFHVQLHNG	NLGGLLSGLQ	SGLVMVSGF	IGLI
			130	140	150	160	170	180
						000	020	240
			190	200	210		230	240
55	orf47a.pe	p GTRI	ISFFTSKRLNV	POIPSPKWVAQ	ASLWLPMLTAM	LMARGVMPWL	SAAFAFAAG	1111
		1111	1111111111	1111111111			11:111111 CAUCACACACAC	(7 T E-17)
	orf47-1	GTRI	ISFFTSKRLNV		ASTMT5WT1W	220	230	240
			190	200	210	220	230	240
<b>CO</b>			250	260	270	280	290	300
60	.e.m	1101717	250 RWWYKPVLKEP	∠OU MT.WTT.FXCVTF	TCI.GI.TAVGAS			
	orf47a.pe	p VQVX.			11111111111	1111111111	111111111	1111
	orf47-1	1111	RWWYKPVLKEP		TGLGLIAVGAS	YFKPAFLNLG	VHLIGVGGI	GVLT
	0114/-1	4041	250	260	270	280	290	300
65					-			
03			310	320	330	340	350	360

35

	orf47a.pep	LGMMARTALGHTG	IPIYPPPKAVI	PVAFWLMMAAT	AVRMVAVES:	SGTAYTHSIR!	rssvlfa
	orf47-1	LGMMARTALGHTGN	PIYPPPKAVI	PVAFWLMMAAT	AVRMVAVES	GTAYTHSIR'	SSVLFA
		310	320	330	340	350	360
5				*			;
-		. 370	380		•		
	orf47a.pep	LALLVYAWKYIPW1	LIRPRSDGRP	ΞX			
	<del>-</del> -	1111111111111		11			
	orf47-1	LALLVYAWKYIPWI	LIRPRSDGRP	GX			
10	. <b>.</b>	370	380				

## Homology with a predicted ORF from N.gonorrhoeae

ORF47 shows 97.1% identity over 172 aa overlap with a predicted ORF (ORF47ng) from N.gonorrhoeae:

```
15
              MKFTKHPVWAMAFRPFYSLAALYGALSVLLWGFGYTGTHELSGFYWHAHEMIWGYAGLVV
                                                            60
       ORF47
               60
       ORF47ng
              MKFTKHPVWAMAFRPFYSLAALYGALSVLLWGFGYTGTHELSGFYWHAHEMIWGYAGLVV
               IAFLLTAVATWTGQPPTRGGVLVGLTIFWLAARIAAFIPGWGASASGILGTLFFWYGAVC
       ORF47
                                                           120
20
               ORF47ng
               IAFLLTAVATWTGQPPTRGGVLVGLTAFWLAARIAAFIPGWGAAASGILGTLFFWYGAVC
                                                           120
                                                           172
       ORF47
               MALPVIRSQNQRNYVAVFALFVLGGTHAAFHVQLHNGNLGGLLSGLQSGLVM
               25
               MALPVIRSQNRRNYVAVFAIFVLGGTHAAFHVQLHNGNLGGLLSGLQSGLVMVWGFIGLI 180
```

The ORF47ng nucleotide sequence <SEQ ID 711> is predicted to encode a protein comprising amino acid sequence <SEQ ID 712>:

```
1 MKFTKHPVWA MAFRPFYSLA ALYGALSVLL WGFGYTGTHE LSGFYWHAHE
51 MIWGYAGLVV IAFLLTAVAT WTGQPPTRGG VLVGLTAFWL AARIAAFIPG
30 101 WGAAASGILG TLFFWYGAVC MALPVIRSON RRNYVAVFAI FVLGGTHAAF
151 HVQLHNGNLG GLLSGLQSGL VMVWGFIGLI GMKIISFFTS KRLKLPQIPS
201 PKWVAHASLW LPMLNAILMA HRVMPWLSAA FPFAAGVIFT VQVYAGGITP
251 IEETSCGSVA GICYRLGNSS G
```

The predicted leader peptide and transmembrane domains are identical (except for an Ile/Ala substitution at residue 87 and an Leu/Ile substitution at position 140) to sequences in the meningococcal protein (see also *Pseudomonas stutzeri* orf396, accession number e246540):

```
TM segments in ORF47ng
                INTEGRAL
                             Likelihood = -5.63
                                                                     52 -
                                                                           68
                                                   Transmembrane
                INTEGRAL
                             Likelihood = -3.88
                                                    Transmembrane
                                                                    169 - 185
                                                                     82 <del>-</del>
40
                INTEGRAL
                             Likelihood = -3.08
                                                    Transmembrane
                                                                          98
                INTEGRAL
                             Likelihood = -1.91
                                                    Transmembrane
                                                                    134 - 150
                                                                    107 - 123
                INTEGRAL
                             Likelihood = -1.44
                                                    Transmembrane
                INTEGRAL
                             Likelihood = -1.38
                                                    Transmembrane
                                                                    227 - 243
```

Further work revealed the complete gonococcal DNA sequence <SEQ ID 713>:

45	1	ATGAAATTTA	CCAAACATCC	CGTCTGGGCA	ATGGCGTTCC	GCCCGTTTTA
	51	TTCACTGGCG	GCACTGTACG	GCGCATTGTC	CGTATTGCTG	TGGGGTTTCG
	. 101	GCTACACGGG	AACGCACGAG	CTGTCCGGTT	TCTATTGGCA	CGCGCATGAG
	151	ATGATTTGGG	GTTATGCCGG	TCTCGTCGTC	ATCGCCTTCC	TGCTGACCGC
	201	CGTCGCCACT	TGGACGGGAC	AGCCGCCCAC	GAGGGGCGGC	GTTCTGGTCG
50	251	GCTTGACCGC	CTTTTGGCTG	GCTGCGCGGA	TTGCCGCCTT	TATCCCGGGT
	301	TGGGGTGCGG	CGGCAAGCGG	CATACTCGGT	ACGCTGTTTT	TCTGGTACGG
	351	CGCGGTGTGC	ATGGCTTTGC	CCGTTATCCG	TtcgCAAAAC	CGGCGCAACT
	401	ATGtcgCCGT	ATTCGCAATA	TTTGTGCTGG	GCGGTACGCA	TGCGgcgTTC
	451	CACGtccAgc	tGCACAACGG	CAACCTAGGC	GGACTCTTGA	GCGGATTGCA
55	501	GTCGGGCCTG	GTTATGGTGT	CGGGCTTTAT	CGGCCTGATT	GGGATGAGGA
	551	TTATTTCGTT	TTTTACGTCC	AAACGGTTGA	ACGTGCCGCA	GATTCCCAGT
	601	CCGAAATGGG	TGGCGCAGGC	TTCGCTGTGG	CTACCCATGC	TGACCGCCAT

5	701 CGGC 751 GTAT 801 CGGZ 851 TCAZ 901 TTGC 951 TCCC	GATGGCG CACGGCGTGA TGCCTTGGCT GTCGGCGGCT TTCGCGTTTG CGGGCGT GATTTTACC GTACAGGTGT ACCGCTGGTG GTATAAACCC ATTGAAAG AACCGATGCT GTGGATTCTG TTTGCCGGCT ATCTGTTTAC ATTGGGG CTGATTCGG TCGGCGCTC CATACGGGCT CGGCGCTCGGT CGGCGCTCGGT CGGCGCTCGGT ATCGATTTA GCCGCCC AAAGCCGTTC CCGTTGCGT TTGGCTGATG ATGGCGCAA CCGTCCG TATGGTTGCC GTATTTTCTT CCGGCACTGC CTACACGCAC CATCCGCA CGTCTTCGGT TTTGTTTGCA CTCGCGCTGC TGGTGTATGC CAAATAC ATTCCGTGC TGATCCGTCC GCGTTCGGAC GGCAGGCCCG CAAATAC ATTCCGTGCC TGATCCGTCC GCGTTCGGAC GGCAGGCCCG CAAAACCCGTAC CTCGCGCTGC TTGGTTATGC CCGCACTGC TTGGTTTGCA CTCGCGCTGC GGCAGGCCCG CGCACGCCCC GCGCACGC GGCAGGCCCCG CGCACGCCCC GCGCACGC GGCAGGCCCC CAAAACCCGTCC TGATCCGTCC GCGTTCGGAC GGCAGGCCCCG CGCACGCCCC GCGCTCCGCC GGCAGGCCCCG CGCGCACGCCCC GCGCTCCC GCGCACGCCCC CAACCGCACGCCCCC GCCACGCCCC GGCAGGCCCCG CAACCGCACGCCCCC TGATCCGCCC GCGCACGCCCC CAACCGCCCCC TTGGTTTTCTTT CCCGCCCCCC CTCGCCCCCC GCCACGCCCCC GCCCCCCCCCC
		in having amino acid sequence <seq 714;="" id="" orf47ng-1="">:</seq>
15 20	1 MKF 51 MIW 101 WGA 151 HVQ 201 PKW 251 VLK	TKHPVWA MAFRPFYSLA ALYGALSVLL WGFGYTGTHE LSGFYWHAHE GYAGLVV IAFLLTAVAT WTGQPPTRGG VLVGLTAFWL AARIAAFIPG AASGILG TLFFWYGAVC MALPVIRSQN RRNYVAVFAI FVLGGTHAAF LHNGNLG GLLSGLQSGL VMVSGFIGLI GMRIISFFTS KRLNVPQIPS VAQASLW LPMLTAILMA HGVMPWLSAA FAFAAGVIFT VQVYRWWYKP EPMLWIL FAGYLFTGLG LIAVGASYFK PAFLNLGVHL IGVGGIGVLT MARTALG HTGNSIYPPP KAVPVAFWLM MAATAVRMVA VFSSGTAYTH ETSSVLFA LALLVYAWKY IPWLIRPRSD GRPG*
	ORF47ng-1 and OR	F47-1 show 97.4% identity in 384 aa overlap:
25	orf47-1.pep	10 20 30 40 50 60 MKFTKHPVWAMAFRPFYSLAALYGALSVLLWGFGYTGTHELSGFYWHAHEMIWGYAGLVV
30	orf47-1.pep orf47ng-1	70 80 90 100 110 120 IAFLLTAVATWTGQPPTRGGVLVGLTIFWLAARIAAFIPGWGASASGILGTLFFWYGAVC
35	orf47-1.pep orf47ng-1	130 140 150 160 170 180 MALPVIRSQNQRNYVAVFALFVLGGTHAAFHVQLHNGNLGGLLSGLQSGLVMVSGFIGLI
40	orf47-1.pep	190 200 210 220 230 240 GTRIISFFTSKRLNVPQIPSPKWVAQASLWLPMLTAMLMAHGVLAWLSAVFAFAAGVIFT
45	orf47ng-1 orf47-1.pep	GMRIISFFTSKRLNVPQIPSPKWVAQASLWLPMLTAILMAHGVMPWLSAAIAIAAGVIFT 190 200 210 220 230 240  250 260 270 280 290 300  WANDENDAY REPORT WILLFAGYLFT GLGLIAV GASYFK PAFLINL GVHLIGV GGIGVLT
50	orf47ng-1	VQVYRWWYKPVLKEPMLWILFAGYLFTGLGLIAVGASYFKPAFLNLGVHLIGVGGIGVLT 250 260 270 280 290 300
55	orf47-1.pep orf47ng-1	310 320 330 340 350 360  LGMMARTALGHTGNPIYPPPKAVPVAFWLMMAATAVRMVAVFSSGTAYTHSIRTSSVLFA
60	orf47-1.pep orf47ng-1	370 380 LALLVYAWKYIPWLIRPRSDGRPGX

Furthermore, ORF47ng-1 shows significant homology to an ORF from Pseudomonas stutzeri:

gnl|PID|e246540 (Z73914) ORF396 protein [Pseudomonas stutzeri] Length = 396 Score = 155 bits (389), Expect = 5e-37

```
Identities = 121/391 (30%), Positives = 169/391 (42%), Gaps = 21/391 (5%)
                    PVWAMAFRPFYSLAALYGALSVLLWGFGYTGTHELSGFY-----WHAHEMIWGYAGLV 59
         Query: 7
                                +LY L++ LW
                    P+W +AFRPF+
                                             +TG
                                                     GF
                                                              WH HEM++G+A
5
         Sbjct: 14 PIWRLAFRPFFLAGSLYALLAIPLWVAAWTGLWP--GFQPTGGWLAWHRHEMLFGFAMAI 71
         Query: 60
                   VIAFLLTAVATWTGQPPTRGGVLVGLTAFWLAARIAAFIPGWGAAASGILGTLFFWYGAV 119
                    V FLLTAV TWTGQ
                                   G LVGL A WLAAR+ ++ G AA
         Sbjct: 72 VAGFLLTAVQTWTGQTAPSGNRLVGLAAVWLAARL-GWLFGLPAAWLAPLDLLFLVALVW 130
10
         MA +
                            + +RNY V + ++ G.
                                                                    +V+ + L
         Sbjct: 131 MMAQMLWAVRQKRNYPIVVVLSLMLGADVLILTGLLQGNDALQRQGVLAGLWLVAALMAL 190
15
         Query: 180 IGMRIISFFTSKRLNVPQIPSP-KWVAQASLWLPMLTAILMAHGV--
                    IG R+I FFT + L
                                  P W+ A L
                                                  + A+L A GV
                                                                 PL FA
         Sbjct: 191 IGGRVIPFFTQRGLGKVDAVKPWVWLDVALLVGTGVIALLHAFGVAMRPQPLLGLLFV-A 249
         Query: 235 AGVIFTVQVYRWWYKPVLKEPMLWILFAGYLFTGLGLIAVGASYF-KPAFXXXXXXXXXX 293
20
                         +++ RW+ K + K +LW L
                                              L+ +
                                                           +F
         Sbjct: 250 IGVGHLLRLMRWYDKGIWKVGLLWSLHVAMLWLVVAAFGLALWHFGLLAQSSPSLHALSV 309
         Query: 294 XXXXXXXXMMARTALGHTGNSIYPPPKAVPVAFWLXXXXXXXXXXXXXXFSSGTAYTHSIR 353
                            M+AR LGHTG + P + AF L
25
          Sbjct: 310 GSMSGLILAMIARVTLGHTGRPLQLPAGIIG-AFVL---FNLGTAARVFLSVAWPVGGLW 365
         Query: 354 TSSVLFALALLVYAWKYIPWLIRPRSDGRPG 384
                     ++V + LA +Y W+Y P L+ R DG PG
          Sbjct: 366 LAAVCWTLAFALYVWRYAPMLVAARVDGHPG 396
30
```

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 85

The following partial DNA sequence was identified in N.meningitidis <SEQ ID 715>:

```
35
                      ..ATGCCGTCTG AAGGTTCAGA CGGCmTCGGT GyCGGGGAAY CAGAAGYGGT
                  51
                         AGCGCATGCC CAATGAGACT TCGTGGGTTT TGAAGCGGGT GTTTTCCAAG
                 101
                         CGTCCCCAGT TGTGGTAACG GTATCCGGTG TCYAArGTCA GCTTGGGYGT
                         GATGTCGAAa CCGACACCGG CGATGACACC AAGACCYAMG CTGCTGATrC
                 151
                         TGTkGCTTTC GTGATAGGSA GGTTTGyTGG kmksAsyTTG TAyrATwkkG
                 201
40
                         CCTssCwsTG kAGmGCCkTk CkyTGGTkkA swGrwArTAG TCGTGGTTTy
                 251
                         TKTTYYCACC GAATGAACYT GATGTTTAAC GTGTCCGTAG GCGACGCGCG
                 301
                        CGCCGATATA GGGTTTGAAT TTATCGTTGA GTTTGAAATC GTAAATGGCG
GACAAGCCGA GAGAAGAAAC GGCGTGGAAG CTGCCGTTTC CCTGATGTTT
                 351
                 401
                         TGTTTGGGTT TCTTTGTAGT TGTTGTTTAT CTCTTCAGTA ACTTTTTTAG
                 451
45
                 501
                         TAGAAGAATT ACTTTCTTTC CATTTTCTGT AACTGGCATA ATCTGCCGCT
                         ATTCTCCAGC CGCCGAAATC ...
```

This corresponds to the amino acid sequence <SEQ ID 716; ORF67>:

```
1 ..MPSEGSDGXG XGEXEXVAHA QXDFVGFEAG VFQASPVVVT VSGVXXQLGX
51 DVETDTGDDT KTXAADXVAF VIGRFXGXXL YXXAXXXXAX XWXXXXSRGF
50 101 XXHRMNLMFN VSVGDARADI GFEFIVEFEI VNGGQAERRN GVEAAVSLMF
151 CLGFFVVVVY LFSNFFSRRI TFFPFSVTGI ICRYSPAAEI ...
```

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from N. gonorrhoeae

ORF67 shows 51.8% identity over 199 aa overlap with a predicted ORF (ORF67ng) from

55 N.gonorrhoeae:

WO 99/24578 PCT/IB98/01665

-409-

	orf67.pep	MPSEGSDGXGXGEXEXVAHA		30 146
	orf67ng	TNFEIAVLSGMTVRVFYCARPAPVNGGRLKMPSEGSDGIGIGESEAVAH 90 100 110 120 130	AQRGEVGEEAG 140	140
5	orf67.pep	VFQASPVVVTVSGVXXQLGXDVETDTGDDTKTXAADXVAFVIGRFXGXX	LYXXAXXXAX	90
	orf67ng		: VRINRNCCVSI	206
	orionig	-		150
10	orf67.pep	XWXXXXSRGFXXHRMNLMFNVSVGDARADIGFEFIVEFEIVNGGQAERR :  :  :: ::      :	1111 11 111	266
	orf67ng	TRVGGKSTCYFFSRIDAVSDVSVGDARTDIGFEFVVEFEIVNGGQAERR	NGVECAVELINE	
	orf67.pep	CLGFFVVVVYLFSNFFSRRITFF-PFSVTGICRYSPAAEI		190
15	orf67ng	::  :  :  :         :    : RLLVFYVKLVAAKSFIILSFQLFYVHGIFIVVPFPVTGIIRGDAPAAEV		326
	Offorma			

The ORF67ng nucleotide sequence <SEQ ID 717> is predicted to encode a protein comprising amino acid sequence <SEQ ID 718>:

```
20 1 MPSETVGSIV NVGVDESVGF SPPFPSIQHF YRFHRIHRIR LFRPPGPMQL
101 VFYCARPAPV NGGRLKMPSE GSDGIGIGES EAVAHAQRGF VGFEAGVFQA
151 SPVVVAVAGV QGQAGRDVYA HARHRAEAQA AAAVAFLIGV FLRMSVRINR
201 NCCVSITRVG GKSTCYFFSR IDAVSDVSVG DARTDIGFEF VVEFEIVNGG
251 QAERRNGVEC AVFLMFRLLV FYVKLVAAKS FIILSFQLFY VHGIFIVVPF
25 301 PVTGIIRGDA PAAEVVADRH PGVDGMRTDV SEIIAYRAYF VFAWSGWFRI
25 351 IVGNAFGGVG *
```

Based on the presence of a several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

# 30 Example 86

The following partial DNA sequence was identified in N.meningitidis <SEQ ID 719>

```
1 ATGTTTGCTT TTTTAGAAGC CTTTTTTGTC GAATACGGTT ATGCGGCTGT

51 TTTTTTTTGTA TTGGTCATCT GCGGTTTCGG CGTGCCGATT CCCGAGGATT

101 TGACCTTGGT AACAGGCGGC GTGATTTCGG GTATGGGTTA TACCAATCCG

151 CATATTATGT TTGCAGTCG TATGCTCGG GTATTGGTCG GGGACGGCAT

201 CATGTTCGCC GCCGGACGAA TTTGGGGGCA GARARTCCTA rGGTTCARAC

251 CTATTGCGSG CATCATGACG CCGRAACGTT ATGAGCAGGT TCAGGAAAAA

301 TTCGACAAAT ACGGTAACTG GGTCTTATTT GTCGCCCGT TCCTGCCCGG

351 TTTGAGAACG GCCGTATTTG TACAGCCGG TATCAGCCGC AAGGTTTCAT

40 401 ACTTGCGTTT TATCATTATG GATGGACTGG CCGCA...
```

This corresponds to the amino acid sequence <SEQ ID 720; ORF78>:

```
1 MFAFLEAFFV EYGYAAVFFV LVICGFGVPI PEDLTLVTGG VISGMGYTNP
51 HIMFAVGMLG VLVGDGIMFA AGRIWGQXXL XFXPIAXIMT PXRYEQVQEK
101 FDKYGNWVLF VARFLPGLRT AVFVTAGISR KVSYLRFIIM DGLAA...
```

# 45 Further work revealed the complete nucleotide sequence <SEQ ID 721>:

50 1 101 151 201 251 301 351 401 55 451	CATATTATGT CATGTTCGCC CTATTGCGCG TTCGACAAAT TTTGAGAACG	TTGGTCATCT AACAGGCGGC TTGCAGTCGG GCCGGACGAA CATCATGACG ACGGTAACTG GCCGTAATTG TATCATTATG	CTTTTTTGTC GCGGTTTCGG GTGATTTCGG TATGCTCGGC TTTGGGGGCA CCGAAACGTT GGTCTTATTT TTACAGCCGG GATGGACTGG ATACGGTGCG	CGTGCCGATT GTATGGGTTA GTATTGGTCG GAAAATCCTA ATGAGCAGGT GTCGCCCGTT TATCAGCCGC CCGCACTGAT	CCCGAGGATT TACCAATCCG GGGACGCAT AGGTTCAAAC TCAGGAAAA TCCTGCCCGG AAGGTTTCAT TTCCGTCCCT
-----------------------------------------------------------------	--------------------------------------------------------------------	-----------------------------------------------------------------------------------------	---------------------------------------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------	---------------------------------------------------------------------------------------

501	GGCGAAAATG	CACAGCCTGC	AATCGGGTAT	TTTTGTTATC	TTGGGTATAG
551	GTGCGACCGT	TGTCGCTTGG	ATTTGGTGGA	AAAAACGCCA	ACGTATCCAG
601	TTTTACCGCA	GCAAATTGAA	AGAAAAGCGG	GCGCAACGCA	AAGCCGCCAA
651	GGCAGCCAAA	AAAGCCGCGC	AAAGCAAACA	ATAA	

5 This corresponds to the amino acid sequence <SEQ ID 722; ORF78-1>:

```
MFAFLEAFFV EYGYAAVFFV LVICGFGVPI PEDLTLVTGG VISGMGYTNP
```

- HIMFAVGMLG VLVGDGIMFA AGRIWGQKIL RFKPIARIMT PKRYEQVQEK 51
- FDKYGNWVLF VARFLPGLRT AVFVTAGISR KVSYLRFIIM DGLAALISVP IWIYLGEYGA HNIDWLMAKM HSLQSGIFVI LGIGATVVAW IWWKKRQRIQ 101
- 151

10 201 FYRSKLKEKR AQRKAAKAAK KAAQSKQ\*

> Computer analysis of this amino acid sequence predicts several transmembrane domains, and also gave the following results:

## Homology with the dedA homologue of H.influenzae (accession number P45280)

ORF78 and the dedA homologue show 58% aa identity in 144aa overlap:

```
15
                     FLEAFFVEYGYAAVFFVLVICGFGVPIPEDLTLVTGGVISGM--GYTNPHIMFAVGMLGV 61
          Orf78: 4
                     FL FF EYGY AV FVL+ICGFGVPIPED+TLV+GGVI+G+
                                                                   N H+M V M+GV
                     FLIGFFTEYGYWAVLFVLIICGFGVPIPEDITLVSGGVIAGLYPENVNSHLMLLVSMIGV 79
          Orf78: 62
                     LVGDGIMFAAGRIWGQXXLXFXPIAXIMTPXRYEQVQEKFDKYGNWVLFVARFLPGLRTA 121
20
                                     L F PI I+T R
                     L GD M+ GRI+G
                                                       V+EKF +YGN VLFVARFLPGLR
          Deda: 80 LAGDSCMYWLGRIYGTKILRFRPIRRIVTLQRLRMVREKFSQYGNRVLFVARFLPGLRAP 139
          Orf78: 122 VFVTAGISRKVSYLRFIIMDGLAA 145
                     +++ +GI+R+VSY+RF+++D AA
25
                 140 IYMVSGITRRVSYVRFVLIDFCAA 163
```

#### Homology with a predicted ORF from N. meningitidis (strain A)

ORF78 shows 93.8% identity over a 145aa overlap with an ORF (ORF78a) from strain A of N. meningitidis:

30	•	10	20	30	40	50	60
	orf78.pep	MFAFLEAFFVEYG <u>Y</u>	AVFFVLVI	CGFGVPIPEDI	TLVTGGVISC	MGYTNPH <u>IM</u> F	AVGMLG
		$\Pi$ HHHHHH $\Pi$		HIIIIIIIII	11111111111	111111111111111111111111111111111111	111111
	orf78a	MFALLEAFFVEYG <u>Y</u>			TLVTGGVISG	MGYTNPH <u>IME</u>	'AVGMLG
25		10	20	30	40	50	60
35							
		70	80	90	100	110	120
	orf78.pep	VLVGDGIMFAAGRIW	VGQXXLXFX	PIAXIMTPXRY	EQVQEKFDKY	GNWVLFVARE	<u>LPGLRT</u>
		1111111111111111				1111111111	111111
40	orf78a	<u>VLVGDGIM</u> FAAGRIW					
40	•	70	80	90	100	110	120
		130	140				
	orf78.pep	AVFVTAGISRKVSYI	LR <u>FIIMDGL</u>	<u> </u>			
4.5			:				
45	orf78a	<u>AVFV</u> TAGISRKVSYI	LR <u>FLIMDGL</u>	<u>AALISVPVWI</u> Y	'LGEYGAHNII	wlmakmhslc	SGIFIA
•		130	140	150	160	170	180

The complete length ORF78a nucleotide sequence <SEQ ID 723> is:

	1	ATGTTTGCCC	TTTTGGAAGC	CTTTTTTGTC	GAATACGGCT	ATGCGGCCGT
	51	GTTTTTCGTT	TTGGTCATCT	GCGGTTTCGG	CGTGCCGATT	CCCGAGGATT
50	101	TGACCTTGGT	AACAGGCGGC	GTGATTTCGG	GTATGGGTTA	TACCAATCCG
	151	CATATTATGT	TTGCAGTCGG	TATGCTCGGC	GTATTGGTCG	GGGACGGCAT
	201	CATGTTCGCC	GCCGGACGCA	TCTGGGGGCA	GAAAATCCTC	AAGTTCAAAC
	251	CGATTGCGCG	CATCATGACG	CCGAAACGTT	ACGCACAGGT	TCAGGAAAAA
	301	TTCGACAAAT	ACGGCAACTG	GGTGTTATTT	GTCGCTCGTT	TCCTGCCCGG
55	351	TTTGCGGACT	GCCGTTTTCG	TTACCGCCGG	CATCAGCCGC	AAAGTATCGT
	401	ATCTGCGCTT	TCTGATTATG	GACGGGCTTG	CCGCGCTGAT	TTCCGTGCCC
	451	GTTTGGATTT	ACTTGGGCGA	GTACGGCGCG	CACAACATCG	ATTGGCTGAT

-411-

```
501 GGCGAAAATG CACAGCCTGC AATCCGGCAT CTTCATCGCA TTGGGCGTGC
    TGGCGGCGGC GCTGGCGTGG TTCTGGTGGC GCAAACGCCG ACATTATCAG
551
601 CTTTACCGCG CACAATTGAG CGAAAAACGC GCCAAACGCA AGGCGGAAAA
651 GGCAGCGAAA AAAGCGGCAC AGAAGCAGCA GTAA
```

This encodes a protein having amino acid sequence <SEQ ID 724>: 5

```
MFALLEAFFV EYGYAAVFFV LVICGFGVPI PEDLTLVTGG VISGMGYTNP
   HIMFAVGMLG VLVGDGIMFA AGRIWGQKIL KFKPIARIMT PKRYAQVQEK
51
```

FDKYGNWVLF VARFLPGLRT AVFVTAGISR KVSYLRFLIM DGLAALISVP 101

VWIYLGEYGA HNIDWLMAKM HSLQSGIFIA LGVLAAALAW FWWRKRRHYQ 151

201 LYRAQLSEKR AKRKAEKAAK KAAQKQQ\*

10

50

ORF78a and ORF78-1 show 89.0% identity in 227 aa overlap:

```
40
                                                     50
       orf78a.pep
                 MFALLEAFFVEYGYAAVFFVLVICGFGVPIPEDLTLVTGGVISGMGYTNPHIMFAVGMLG
                  15
       orf78-1
                 MFAFLEAFFVEYGYAAVFFVLVICGFGVPIPEDLTLVTGGVISGMGYTNPHIMFAVGMLG
                       10
                               20
                                      30
                                              40
                                                     50
                                      90
                                             100
                 VLVGDGIMFAAGRIWGQKILKFKPIARIMTPKRYAQVQEKFDKYGNWVLFVARFLPGLRT
       orf78a.pep
                  20
        orf78-1
                 VLVGDGIMFAAGRIWGQKILRFKPIARIMTPKRYEQVQEKFDKYGNWVLFVARFLPGLRT
                               80
                                      90
                                             100
                                                          1 120
                                                    110
                       130
                              140
                                     150
                                             160
25
        orf78a.pep
                 AVFVTAGISRKVSYLRFLIMDGLAALISVPVWIYLGEYGAHNIDWLMAKMHSLQSGIFIA
                  orf78-1
                 AVFVTAGISRKVSYLRFIIMDGLAALISVPIWIYLGEYGAHNIDWLMAKMHSLQSGIFVI
                              140
                                     150
                                             160
30
                       190
                              200
                                     210
                                             220
                  LGVLAAALAWFWWRKRRHYQLYRAQLSEKRAKRKAEKAAKKAAQKQQX
        orf78a.pep
                 orf78-1
                       190
                              200
35
```

#### Homology with a predicted ORF from N.gonorrhoeae

ORF78 shows 97.4% identity over 38 aa overlap with a predicted ORF (ORF78ng) from N. gonorrhoeae:

```
orf78.pep
                     XXLXFXPIAXIMTPXRYEQVQEKFDKYGNWVLFVARFLPGLRTAVFVTAGISRKVSYLRF
40
                                               orf78ng
                                              YPVLFVARFLPGLRTAVFVTAGISRKVSYLRF
         orf78.pep
                     IIMDGLAA
                                                                            145
                     :1111111
45
                     LIMDGLAALISVPVWIYLGEYGAHNIDWLMAKMHSLQSGIFIALGVLAAALAWFWWRKRR
```

The ORF78ng nucleotide sequence <SEQ ID 725> is predicted to encode a protein comprising amino acid sequence <SEQ ID 726>:

```
..YPVLFVARFL PGLRTAVFVT AGISRKVSYL RFLIMDGLAA LISVPVWIYL
```

51 GEYGAHNIDW LMAKMHSLQS GIFIALGVLA AALAWFWWRK RRHYQLYRAQ

101 LSEKRAKRKA EKAAKKAAOK OO\*

Further work revealed the complete gonococcal nucleotide sequence <SEQ ID 727>:

```
atgtttgccc tttTggaagc CTTTTTTGTC GAAtacggCt atgcGGCCGT
                    GTTTTCGTT TTGGTCATCT GCGGTTTCGG CGTGCCGATT CCCGAAGATT
                51
                    TGACCTTGGT AACGGGCGGC GTGATTTCGG GTATGGGTTA TACCAATCCG
               101
55
                    CATATTATGT TTGCGGTCGG TATGCTCGGC GTGTTGGCGG GCGACGGCGT
               151
                    GATGTTTGCC GCCGGACGCA TCTGGGGGCA GAAAATCCTC AAGTTCAAAC
               201
                    CGATTGCGCG CATCATGACG CCGAAACGTT ACGCGCAGGT TCAGGAAAAA
               251
               301
                    TTCGACAAAT ACGGCAACTG GGTTCTGTTT GTCGCCCGTT TCCTGCCGGG
```

WO 99/24578 PCT/IB98/01665

-412-

351

60

```
TTTGCGGACT GCCGTTTTCG TTACCGCCGG CATCAGCCGC AAAGTATCGT
                   ATCTGCGCTT TCTGATTATG GACGGGCTGG CCGCGCTGAT TTCCGTGCCC
               401
                   GTTTGGATTT ACTTGGGCGA GTACGGCGCG CACAACATCG ATTGGCTGAT
               451
                   GGCGAAAATG CACAGCCTGC AATCGGGCAT CTTCATCGCA TTGGGCGTGC
               501
5
                   TGGCGGCGGC GCTGGCGTGG TTCTGGTGGC GCAAACGCCG ACATTATCAG
               551
                   CTTTACCGCG CACAATTGAG CGAAAAACGC GCCAAACGCA AGGCGGAAAA
               601
                   GGCAGCGAAA AAAGCGGCAC AGAAGCAGCA GTAa
               651
     This corresponds to the amino acid sequence <SEQ ID 728; ORF78ng-1>:
                   MFALLEAFFV EYGYAAVFFV LVICGFGVPI PEDLTLVTGG VISGMGYTNP HIMFAVGMLG VLAGDGVMFA AGRIWGQKIL KFKPIARIMT PKRYAQVQEK
10
                   FDKYGNWVLF VARFLPGLRT AVFVTAGISR KVSYLRFLIM DGLAALISVP
               101
               151
                   VWIYLGEYGA HNIDWLMAKM HSLQSGIFIA LGVLAAALAW FWWRKRRHYQ
                   LYRAQLSEKR AKRKAEKAAK KAAQKQQ*
               201
     ORF78ng-1 and ORF78-1 show 88.1% identity in 227 aa overlap:
15
                                       20
                      MFAFLEAFFVEYGYAAVFFVLVICGFGVPIPEDLTLVTGGVISGMGYTNPHIMFAVGMLG
          orf78-1.pep
                       orf78ng-1
                      MFALLEAFFVEYGYAAVFFVLVICGFGVPIPEDLTLVTGGVISGMGYTNPHIMFAVGMLG
                                       20
                                                 30
                                                          40
                                                                    50
                                                                              60
20
                              70
                                       80
                                                 90
                                                         100
                                                                   110
                                                                             120
                      VLVGDGIMFAAGRIWGQKILRFKPIARIMTPKRYEQVQEKFDKYGNWVLFVARFLPGLRT
          orf78-1.pep
                       orf78ng-1
                      VLAGDGVMFAAGRIWGQKILKFKPIARIMTPKRYAQVQEKFDKYGNWVLFVARFLPGLRT
25
                              70
                                       80
                                                 90
                                                         100
                                                                   110
                                                                             120
                                                150
                                                          160
                      AVFVTAGISRKVSYLRFIIMDGLAALISVPIWIYLGEYGAHNIDWLMAKMHSLQSGIFVI
          orf78-1.pep
                       30
          orf78ng-1
                       AVFVTAGISRKVSYLRFLIMDGLAALISVPVWIYLGEYGAHNIDWLMAKMHSLQSGIFIA
                             130
                                       140
                                                150
                                                         160
                                                                   170
                                                                             180
                             190
                                       200
                                                210
                                                          220
                      LGIGATVVAWIWWKKRQRIQFYRSKLKEKRAQRKAAKAAKKAAQSKQX
          orf78-1.pep
35
                       orf78ng-1
                       LGVLAAALAWFWWRKRRHYQLYRAQLSEKRAKRKAEKAAKKAAQKQQX
                             190
                                       200
                                                210
                                                          220
     Furthermore, orf78ng-1 shows homology to the dedA protein from H.influenzae:
          sp|P45280|YG29_HAEIN HYPOTHETICAL PROTEIN HI1629 >gi|1073983|pir||D64133 dedA
40
          protein (dedA) homolog - Haemophilus influenzae (strain Rd KW20)
          >gi|1574476 (U32836) dedA protein (dedA) [Haemophilus influenzae] Length = 212
           Score = 223 bits (563), Expect = 7e-58
           Identities = 108/182 (59%), Positives = 140/182 (76%), Gaps = 2/182 (1%)
45
          Query: 5
                     LEAFFVEYGYAAVFFVLVICGFGVPIPEDLTLVTGGVISGM--GYTNPHIMFAVGMLGVL 62
                     L FF EYGY AV FVL+ICGFGVPIPED+TLV+GGVI+G+
                                                                N H+M V M+GVL
          Sbict: 21
                    LIGFFTEYGYWAVLFVLIICGFGVPIPEDITLVSGGVIAGLYPENVNSHLMLLVSMIGVL 80
          Query: 63
                     AGDGVMFAAGRIWGQKILKFKPIARIMTPKRYAQVQEKFDKYGNWVLFVARFLPGLRTAV 122
50
                     AGD M+ GRI+G KIL+F+PI RI+T +R V+EKF +YGN VLFVARFLPGLR +
          Sbjct: 81
                     AGDSCMYWLGRIYGTKILRFRPIRRIVTLQRLRMVREKFSQYGNRVLFVARFLPGLRAPI 140
          Query: 123 FVTAGISRKVSYLRFLIMDGLAALISVPVWIYLGEYGAHNIDWLMAKMHSLQSGIFIALG 182
                     ++ +GI+R+VSY+RF+++D AA+ISVP+WIYLGE GA N+DWL ++
 55
           Sbjct: 141 YMVSGITRRVSYVRFVLIDFCAAIISVPIWIYLGELGAKNLDWLHTQIQKGQIVIYIFIG 200
           Query: 183 VL 184
           Sbjct: 201 YL 202
```

WO 99/24578

-413-

Based on this analysis, including the presence of putative transmembrane domains, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 87

5 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 729>:

```
ATGAAAAAT TATTGGCGGC CGTGATGATG GCAGGTTTGG CAGGCGCGGT
                51
                    TTCCGCCGCC GGAGTCCACG TTGAGGACGG CTGGGCGCGC ACCACCGTCG
                    AAGGTATGAA AATAGGCGGC GCGTTCATGA AAATCCACAA CGACGAAGCC
               101
               151
                    AAACAAGACT TTTTGCTCGG CGGAAGCAGC CCCGTTGCCG ACCGCGTCGA
10
               201 AGTGCATACC CACATCAACG ACAACGGCGT GATGCGGATG CGCGAAGTCG
                    AAGGCGGCGT GCCTTTGGAA GCGAAATCCG TTACCGAACT CAAACCCGGC
               251
               301
                    AGCTATCATG TGATGTTTAT GGGTTTGAAA AAACAATTAA AAGAGGGCGA
               351
                    TAAAATTCCC GTTACCCTGA AATTTAAAAA CGCCAAAGCG CAAACCGTCC
               401 AACTGGAAGT CAAAATCGCG CCGATGCCGG CAATGAACCA C...
```

15 This corresponds to the amino acid sequence <SEQ ID 730; ORF79>:

```
1 MKKLLAAVMM AGLAGAVSAA GVHVEDGWAR TTVEGMKIGG AFMKIHNDEA
51 KQDFLLGGSS PVADRVEVHT HINDNGVMRM REVEGGVPLE AKSVTELKPG
101 SYHVMFMGLK KQLKEGDKIP VTLKFKNAKA QTVQLEVKIA PMPAMNH..
```

Further work revealed the complete nucleotide sequence <SEQ ID 731>:

```
20
                    ATGAAAAAT TATTGGCGGC CGTGATGATG GCAGGTTTGG CAGGCGCGGT
                    TTCCGCCGCC GGAGTCCACG TTGAGGACGG CTGGGCGCGC ACCACCGTCG
                101
                    AAGGTATGAA AATAGGCGGC GCGTTCATGA AAATCCACAA CGACGAAGCC
                151
                    AAACAAGACT TTTTGCTCGG CGGAAGCAGC CCCGTTGCCG ACCGCGTCGA
                    AGTGCATACC CACATCAACG ACAACGGCGT GATGCGGATG CGCGAAGTCG
25
                    AAGGCGGCGT GCCTTTGGAA GCGAAATCCG TTACCGAACT CAAACCCGGC
                251
                    AGCTATCATG TGATGTTTAT GGGTTTGAAA AAACAATTAA AAGAGGGCGA
                    TAAAATTCCC GTTACCCTGA AATTTAAAAA CGCCAAAGCG CAAACCGTCC
                351
                    AACTGGAAGT CAAAATCGCG CCGATGCCGG CAATGAACCA CGGTCATCAC
                    CACGGCGAAG CGCATCAGCA CTAA
```

This corresponds to the amino acid sequence <SEQ ID 732; ORF79-1>:

```
1 MKKLLAAVMM AGLAGAVSAA GVHVEDGWAR TTVEGMKIGG AFMKIHNDEA
```

151 HGEAHQH\*

35 Computer analysis of this amino acid sequence revealed a putative leader peptide and also gave the following results:

Homology with a predicted ORF from N.meningitidis (strain A)

ORF79 shows 94.6% identity over a 147aa overlap with an ORF (ORF79a) from strain A of N. meningitidis:

40		10	20	30	40	50	60
	orf79.pep	MKKLLAAVMMAG	<u>Laga</u> vsaagvi	IVEDGWARTTV.	EGMKIGGAFM	KIHNDEAKQD!	FLLGGSS
		11 111111111			1111:1111		
	orf79a	MKXLLAAVMMAG	<u>Laga</u> vsaagii	IVEDGWARTTV	EGMKMGGAFMI	KIHNDEAKQD	FLLGGSS
		10	20	30	40	50	60
45							
		70	80	90	100	110	120
	orf79.pep	PVADRVEVHTHI	ndngvmrmrev			_	KEGDKIP
		111111111111			111111111		1 11111
	orf79a	PVADRVEVHTHI	ndngvmrmrev	/EGGVPLEAKS	VTELKPGSYH	VMFMGXKKQL	KXGDKIP
50		70	80	90	100	110	120

<sup>51</sup> KQDFLLGGSS PVADRVEVHT HINDNGVMRM REVEGGVPLE AKSVTELKPG

<sup>101</sup> SYHVMFMGLK KQLKEGDKIP VTLKFKNAKA QTVQLEVKIA PMPAMNHGHH

	orf79.pep	130 140 VTLKFKNAKAQTVQLEVKIAPMPAMNH
5	orf79a	
	The complete lea	ngth ORF79a nucleotide sequence <seq 733="" id=""> is:</seq>
	1 51	ATGAAANAAC TATTGGCAGC CGTGATGATG GCAGGTTTGG CAGGCGCGGT TTCCGCCGCC GGAATCCACG TTGAGGACGG CTGGGCGCGC ACCACCGTCG
10	101 151	AAGGTATGAA AATGGGCGGC GCGTTCATGA AAATCCACAA CGACGAAGCC AAACAAGACT TTTTGCTCGG CGGAAGCAGC CCTGTTGCCG ACCGCGTCGA
	201	AGTGCATACC CATATCAATG ATAACGGTGT GATGCGGATG CGCGAAGTCG AAGGCGGCGT GCCTTTGGAG GCGAAATCCG TTACCGAACT CAAACCCGGC
15	301	AGCTATCATG TCATGTTTAT GGGTNTGAAA AAACAATTAA AAGANGGCGA CAAGATTCCC GTTACCCTGA AATTTAAAAA CGCCAAAGCA CAAACCGTCC
13	401 451	AACTGGAAGT CAAAACCGCG CCGATGTCGG CAATGGACCA CGGTCATCAC CACGGCGAAG CGCATCAGCA CTAA
		protein having amino acid sequence <seq 734="" id="">:</seq>
	1	
20	51 101	
		SYHVMFMGXK KQLKXGDKIP VTLKFKNAKA QTVQLEVKTA PMSAMDHGHH HGEAHQH*
	ORF79a and OR	F79-1 show 94.9% identity in 157 aa overlap:
25	orf79a.pe	10 20 30 40 50 60  MKXLLAAVMMAGLAGAVSAAGIHVEDGWARTTVEGMKMGGAFMKIHNDEAKQDFLLGGSS
	orf79-1	
	011/3 1	10 20 30 40 50 60
30	orf79a.pej	70 80 90 100 110 120 PVADRVEVHTHINDNGVMRMREVEGGVPLEAKSVTELKPGSYHVMFMGXKKQLKXGDKIP
	orf79-1	
35	<b>V</b>	70 80 90 100 110 120
	orf79a.pe	130 140 150 P VTLKFKNAKAQTVQLEVKTAPMSAMDHGHHHGEAHQHX
	orf79-1	
40		130 140 150
	Homology with	a predicted ORF from N. gonorrhoeae
	ORF79 shows !	96.1% identity over 76 aa overlap with a predicted ORF (ORF79ng) from
	N.gonorrhoeae:	
45	orf79.pep	FMKIHNDEAKQDFLLGGSSPVADRVEVHTHINDNGVMRMREVEGGVPLEAKSVTELKPGS 101
	orf79ng	INDNGVMRMREVKGGVPLEAKSVTELKPGS 30
50	orf79.pep	YHVMFMGLKKQLKEGDKIPVTLKFKNAKAQTVQLEVKIAPMPAMNH 147
	orf79ng	YHVMFMGLKKQLKEGDKIPVTLKFKNAKAQTVQLEVKTAPMSAMNHGHHHGEAHQH 86
		acleotide sequence <seq 735="" id=""> was predicted to encode a protein comprising</seq>
	amino acid sequ	ence <seq 736="" id="">:</seq>
55	1 51	INDNGVMRMR EVKGGVPLEA KSVTELKPGS YHVMFMGLKK QLKEGDKIPV TLKFKNAKAQ TVQLEVKTAP MSAMNHGHHH GEAHQH*

Further work revealed the complete gonococcal DNA sequence <SEQ ID 737>:

-415-

```
1 ATGAAAAAT TATTGGCAGC CGTGATGATG GCAGGTTTGG CAGGCGCGGT
                51
                   TTccgccgCc GGagTccAtG TCGAggACGG CTGGGCGCGc accaCTGtcg
                   aaggtATgaa aatggGCGGC GCgttCATga aaATCCACAA CGACGaaGcc
               101
                   atacaaGACt ttgtgcTCgg CGGaagcatg cccgttgccg accgcGTCGA
               151
               201 AGTGCAtaca cacATCAACG ACAACGGCGT GATGCGTATG CGCGAAGTCA
5
               251
                   AAGGCGGCGT GCCTTTGGAG GCGAAATCCG TTACCGAACT CAAACCCGGC
                   AGCTATCACG TGATGTTTAT GGGTTTGAAA AAACAACTGA AAGAGGGCGA
               301
               351
                   CAAGATTCCC GTTACCCTGA AATTTAAAAA CGCCAAAGCG CAAACCGTCC
               401
                   AACTGGAAGT CAAAACCGCG CCGATGTCGG CAATGAACCA CGGTCATCAC
10
               451
                   CACGGCGAAG CGCATCAGCA CTAA
     This corresponds to the amino acid sequence <SEO ID 738; ORF79ng-1>:
                   MKKLLAAVMM AGLAGAVSAA GVHVEDGWAR TTVEGMKMGG AFMKIHNDEA
                   IQDFVLGGSM PVADRVEVHT HINDNGVMRM REVKGGVPLE AKSVTELKPG
                51
               101
                   SYHVMFMGLK KQLKEGDKIP VTLKFKNAKA QTVQLEVKTA PMSAMNHGHH
15
               151
                   HGEAHQH*
     ORF79ng-1 and ORF79-1 show 95.5% identity in 157 aa overlap:
                                                                    50
                      {\tt MKKLLAAVMMAGLAGAVSAAGVHVEDGWARTTVEGMKIGGAFMKIHNDEAKQDFLLGGSS}
          orf79-1.pep
                      20
                      MKKLLAAVMMAGLAGAVSAAGVHVEDGWARTTVEGMKMGGAFMKIHNDEAIQDFVLGGSM
          orf79ng-1
                              10
                                       20
                                                 30
                                                          40
                                                                   50
                                       80
                                                 90
                                                         100
                                                                  110
                                                                            120
          orf79-1.pep PVADRVEVHTHINDNGVMRMREVEGGVPLEAKSVTELKPGSYHVMFMGLKKOLKEGDKIP
25
                      orf79ng-1
                      PVADRVEVHTHINDNGVMRMREVKGGVPLEAKSVTELKPGSYHVMFMGLKKQLKEGDKIP
                              70
                                       80
                                                 90
                                                         100
                                                                  110
                                                                            120
                                      140
30
          orf79-1.pep
                      VTLKFKNAKAQTVQLEVKIAPMPAMNHGHHHGEAHQHX
                      VTLKFKNAKAQTVQLEVKTAPMSAMNHGHHHGEAHQHX
          orf79ng-1
                             130
                                      140
     Furthermore, ORF79ng-1 shows significant homology to a protein from Aquifex aeolicus:
35
          gi|2983695 (AE000731) putative protein [Aquifex aeolicus] Length = 151
           Score = 63.6 bits (152), Expect = 6e-10
           Identities = 38/114 (33%), Positives = 58/114 (50%), Gaps = 1/114 (0%)
          Query: 24 VEDGWARTTVEGMKMGGAFMKIHNDEAIQDFVLGGSMPVADRVEVHTHINDNGVMRMREV 83
40
                                      M I N+
                                                D+++G
                                                        +A RVE+H + +N V +M
                    VKHPWVMEPPPGPNTTMMGMIIVNEGDEPDYLIGAKTDIAQRVELHKTVIENDVAKMVPQ 86
          Sbjct: 27
                    KGGVPLEAKSVTELKPGSYHVMFMGLKKQLKEGDKIPVTLKFKNAKAQTVQLEV 137
          Query: 84
                     + + + K
                               ΕK
                                     YHVM +GLKK++KEGDK+ V L F+ +
```

Based on this analysis, it is predicted that the proteins from N.meningitidis and N.gonorrhoeae, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Sbjct: 87 ER-IEIPPKGKVEFKHHGYHVMIIGLKKRIKEGDKVKVELIFEKSGKITVEAPV 139

45

50

TV+

ORF79-1 (15.6kDa) was cloned in the pET vector and expressed in E.coli, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 18A shows the results of affinity purification of the His-fusion protein. Purified His-fusion protein was used to immunise mice, whose sera were used for ELISA (positive result) and FACS analysis (Figure 18B) These experiments confirm that ORF79-1 is a surface-exposed protein, and that it is a useful immunogen.

#### Example 88

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 739>:

```
ATGACGGTAA CTGCGGCCGA AGGCGGCAAA GCTGCCAAGG CGTTAAAAAA
 5
                     ATATCTGATT ACGGGCATTT TGGTCTGGCT GCCGATTGCG GTAACGGTTT
                51
               101
                     GGGTGGTTTC CTATATCGTT TCCGCGTCCG ATCAGCTCGT CAACCTGCTG
               151
                     CCGAAGCAAT GGCGGCCGCA ATATGTTTTG GGGTTTAATA TCCCGGGGCT
                     GGGCGTTATC GTTGCCATTG CCGTATTGTT TGTAACCGGA TTGTTTGCCG
               201
                     CCAACGTATT GGGTCGGCAG ATCCTCGCCG CGTGGGACAG CCTGTTGGGG
               251
10
                     CGGATTCCGG TTGTGAAAtC CATCTATTCG AGTGTGAAAA AAGTATCCGA
               301
               351
                     ATacgTGCTG TCCGACAGCA GCCGTTCGTT TAAAACGCCG GTACTCGTGC
                     CGTTTCCCCA GCCCGGTATT TGGACGATYG CTTTCGTGTC AGGGCAGGTG
               401
               451
                     TCGAATGCGG TTAAGGCCGC ATTGCCGAAs GACGGCGATT ATCTTTCCGT
                     GTATGTTCCG ACCACGCCGA ATCCGACCGG CGGTTACTAT ATTATGGTAA
               501
15
               551
                     AGAAAAGCGA TGTGCGCGAA CTCGATATGA GCGTGGACGA ASCATTGAAA
                     TATGTGATTT CGCTGGGTAT GGTCATCCCT GACGACCTGC CCGTCAAAAC
                601
                651
                     ATTGGCASGA CCTATGCCGT CTGAAAAGGC GGATTTGCCC GAACAACAAT
                701
```

This corresponds to the amino acid sequence <SEQ ID 740; ORF98>:

```
20 1 MTVTAAEGGK AAKALKKYLI TGILVWLPIA VTVWVVSYIV SASDQLVNLL
51 PKQWRPQYVL GFNIPGLGVI VAIAVLFVTG LFAANVLGRQ ILAAWDSLLG
101 RIPVVKSIYS SVKKVSEYVL SDSSRSFKTP VLVPFPQPGI WTIAFVSGQV
151 SNAVKAALPX DGDYLSVYVP TTPNPTGGYY IMVKKSDVRE LDMSVDEXLK
201 YVISLGMVIP DDLPVKTLAX PMPSEKADLP EQO*
```

25 Further work revealed the complete nucleotide sequence <SEQ ID 741>:

	1	ATGACGGAAC	nTGCGGCCGA	AGGCGGCAAA	GCTGCCAArG	CGTTAAAAAA
	51	ATATCTGATT	ACGGGCATTT	TGGTCTGGCT	GCCGATTGCG	GTAACGGTTT
	101	GGGTGGTTTC	CTATATCGTT	TCCGCGTCCG	ATCAGCTCGT	CAACCTGCTG
	151	CCGAAGCAAT	GGCGGCCGCA	ATATGTTTTG	GGGTTTAATA	TCCCGGGGCT
30	201	GGGCGTTATC	GTTGCCATTG	CCGTATTGTT	TGTAACCGGA	TTGTTTGCCG
	251	CCAACGTATT	GGGTCGGCAG	ATCCTCGCCG	CGTGGGACAG	CCTGTTGGGG
,	301	CGGATTCCGG	TTGTGAAATC	CATCTATTCG	AGTGTGAAAA	AAGTATCCGA
	351	ATCGCTGCTG	TCCGACAGCA	GCCGTTCGTT	TAAAACGCCG	GTACTCGTGC
	401	CGTTTCCCCA	GCCCGGTATT	TGGACGATTG	CTTTCGTGTC	AGGGCAGGTG
35	451	TCGAATGCGG	TTAAGGCCGC	ATTGCCGAAG	GACGGCGATT	ATCTTTCCGT
	501	GTATGTTCCG	ACCACGCCGA	ATCCGACCGG	CGGTTACTAT	ATTATGGTAA
	551	AGAAAAGCGA	TGTGCGCGAA	CTCGATATGA	GCGTGGACGA	AGCATTGAAA
	601	TATGTGATTT	CGCTGGGTAT	GGTCATCCCT	GACGACCTGC	CCGTCAAAAC
	651	ATTGGCAGGA	CCTATGCCGT	CTGAAAAGGC	GGATTTGCCC	GAACAACAAT
40	701	AA				

This corresponds to the amino acid sequence <SEQ ID 742; ORF98-1>:

```
45 MTEXAAEGGK AAKALKKYLI TGILVWLPIA VTVWVVSYIV SASDQLVNLL
51 PKQWRPQYVL GFNIPGLGVI VAIAVLFVTG LFAANVLGRQ ILAAWDSLLG
101 RIPVVKSIYS SVKKVSESLL SDSSRSFKTP VLVPFPQPGI WTIAFVSGQV
151 SNAVKAALPK DGDYLSVYVP TTPNPTGGYY IMVKKSDVRE LDMSVDEALK
201 YVISLGMVIP DDLPVKTLAG PMPSEKADLP EQQ*
```

Computer analysis of this amino acid sequence gave the following results:

## Homology with a predicted ORF from N. meningitidis (strain A)

ORF98 shows 96.1% identity over a 233aa overlap with an ORF (ORF98a) from strain A of N.

50 meningitidis:

5	orf98.pep orf98a	70 80 90 100 110 120 GFNIPGLGVIVAIAVLFVTGLFAANVLGRQILAAWDSLLGRIPVVKSIYSSVKKVSEYVL !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
10	orf98.pep orf98a	130 140 150 160 170 180 SDSSRSFKTPVLVPFPQPGIWTIAFVSGQVSNAVKAALPXDGDYLSVYVPTTPNPTGGYY
15	orf98.pep orf98a	190 200 210 220 230 IMVKKSDVRELDMSVDEXLKYVISLGMVIPDDLPVKTLAXPMPSEKADLPEQQX
	The complete lea	ngth ORF98a nucleotide sequence <seq 743="" id=""> is:</seq>
20	1 51 101	ATGACGGAAC CTGCGGCCGA AGGCGGCAAA GCTGCCAAGG CGTTAAAAAA ATATCTGATT ACGGGCATTT TGGTCTGGCT GCCGATTGCG GTAACGGTTT GGGTGGTTTC CTATATCGTT TCCGCGTCCG ATCAGCTCGT CAACCTGCTG
25	151 201 251 301 351	CCGAAGCAAT GGCGGCCGCA ATATGTTTTG GGGTTTAATA TCCCGGGGCT GGGCGTTATC GTTGCCATTG CCGTATTGTT TGTAACCGGA TTATTTGCCG CAAACGTATT GGGCCGGCAG ATTCTTGCCG CGTGGGACAG CTTGTTGGGG CGGATTCCGG TTGTGAAGTC CATCTATTCG AGTGTGAAAA AAGTATCCGA NTCGTTGCTG TCCGACAGCA GCCGTTCGTT TAAAACACCA GTACTCGTGC
30	401 451 501 551 601 651	CGTTTCCCCA ATCGGGTATT TGGACAATCG CATTCGTGTC CGGTCAGGTG TCGAATGCGG TTAAGGCCGC ATTGCCGAAG GACGGCGATT ATCTTTCCGT GTATGTTCCG ACCACGCCGA ATCCGACCGG CGGTTACTAT ATTATGGTAA AGAAAAGCGA TGTGCGCGAA CTCGATATGA GCGTGGACGA AGCGTTGAAA TATGTGATTT CGCTGGGTAT GGTCATCCCT GACGACCTGC CCGTCAAAAC ATTGGCAGGA CCTATGCCGT CTGAAAAGGC GGATTTGCCC GAACAACAAT
35	701	protein having amino acid sequence <seq 744="" id="">:</seq>
40	1 51 101 151 201	MTEPAAEGGK AAKALKKYLI TGILVWLPIA VTVWVVSYIV SASDQLVNLL PKQWRPQYVL GFNIPGLGVI VAIAVLEVTG LEAANVLGRQ ILAAWDSLLG RIPVVKSIYS SVKKVSXSLL SDSSRSFKTP VLVPFPQSGI WTIAFVSGQV SNAVKAALPK DGDYLSVYVP TTPNPTGGYY IMVKKSDVRE LDMSVDEALK YVISLGMVIP DDLPVKTLAG PMPSEKADLP EQQ*
	ORF98a and OR	EF98-1 show 98.7% identity in 233 aa overlap:
45	orf98a.pe	10 20 30 40 50 60  MTEPAAEGGKAAKALKKYLITGILVWLPIAVTVWVVSYIVSASDQLVNLLPKQWRPQYVL
50	orf98a.pe	70 80 90 100 110 120  GFNIPGLGVIVAIAVLFVTGLFAANVLGRQILAAWDSLLGRIPVVKSIYSSVKKVSXSLL
55	orf98a.pe	130 140 150 160 170 180 p SDSSRSFKTPVLVPFPQSGIWTIAFVSGQVSNAVKAALPKDGDYLSVYVPTTPNPTGGYY
60	orf98a.pe	130 140 150 160 170 180  190 200 210 220 230  p IMVKKSDVRELDMSVDEALKYVISLGMVIPDDLPVKTLAGPMPSEKADLPEQQX
65		190 200 210 220 230

### Homology with a predicted ORF from N. gonorrhoeae

ORF98 shows 95.3% identity over a 233 aa overlap with a predicted ORF (ORF98ng) from *N.gonorrhoeae*:

	•	10	20	30	40	50	60 .	
5	orf98.pep	MTVTAAEGGKAAK	LKKYLITGIL	VWLPIAVTVW	VVSYIVSASDQ	LVNLLPKQW	RPQYVL	60
		11 11111111		111111111	HIIIIIIIII		111111	
	orf98ng	MTEPAAEGGKAAK	LKKYLITGIL	VWLPIAVTVW	VVSYIVSASDQ	LVNLLPKQW	RPQYVL	60
	orf98.pep	GFNIPGLGVIVAI	VLFVTGLFAAI	NVLGRQILAA	WDSLLGRIPVV	KSIYSSVKK	VSEYVL	120
10				1111111	HILL HILL		111 :1	
	orf98ng	GFNIPGLGVIVAI	VLFVTGLFAA	NVLGRQILAA	WDSLLXRIPVV	KSIYSSVKK	VSESLL	120
	orf98.pep	SDSSRSFKTPVLV	PFPQPGIWTIA	FVSGQVSNAV	KAALPXDGDYL	SVYVPTTPN	PTGGYY	180
				1111111111	11111 11111	11111111	111111	
15	orf98ng	SDSSRSFKTPVLV	PFPQSGIWTIA	FVSGQVSNAV	KAALPQDGDYL	SVYVPTTPN	PTGGYY	180
	orf98.pep	IMVKKSDVRELDM:	SVDEXLKYVIS:	LGMVIPDDLP	VKTLAXPMPSE	KADLPEQQ	233	
					11111 111 1	11:1111		
	orf98ng	IMVKKSDVRELDM					233	
20	The complete lengt	h ODEOgna mode	otida saguar	oo <ceo ii<="" td=""><td>715&gt; in</td><td>a: _</td><td></td><td></td></ceo>	715> in	a: _		

The complete length ORF98ng nucleotide sequence <SEQ ID 745> is predicted to encode a protein having amino acid sequence <SEQ ID 746>:

```
1 MTEPAAEGGK AAKALKKYL<u>I TGILVWLPIA VTVWVVSYI</u>V SASDQLVNLL
51 PKQWRPQYVL GFNIPGLGV<u>I VAIAVLFVTG LFAAN</u>VLGRQ ILAAWDSLLX
101 RIPVVKSIYS SVKKVSESLL SDSSRSFKTP VLVPFPQSGI WTIAFVSGQV
25 151 SNAVKAALPQ DGDYLSVYVP TTPNPTGGYY IMVKKSDVRE LDMSVDEALK
201 YVISLGMVIP DDLPVKTLAG PMPPEKAELP EQQ*
```

## Further work revealed the complete nucleotide sequence <SEQ ID 747>:

	1	ATGACGGAAC	CTGCGGCCGA	AGGCGGCAAA	GCTGCCAAGG	CGTTAAAAAA
	51	ATATCTGATT	ACAGGCATTT	TGGTCTGGCT	GCCGATTGCG	GTAACGGTTT
30	101	GGGTGGTTTC	CTATATCGTT	TCCGCGTCCG	ACCAGCTTGT	CAACCTGCTG
	151	CCGAAGCAAT	GGCGGCCGCA	ATATGTTTTG	GGGTTTAATA	TCCCCGGGCT
	201	CGGCGTTATT	GTTGCCATTG	CCGTATTGTT	TGTAACCGGA	TTATTTGCCG
	251	CAAACGTGTT	GGGCCGGCAG	ATTCTTGCCG	CGTGGGACAG	CCTGTTqqqq
35	301	cggaTTCCGG	TTGTCAAATC	CATCTATTCG	AGTGTGAAAA	AAGTATCCGA
	351			GCCGTTCGTT		
	401	CGTTTCCCCA	ATCGGGTATT	TGGACAATCG	CATTCGTGTC	CGGTCAGGTG
	451			ATTGCCGCAG		
	501	GTATGTCCCG	ACCACGCCCA	ACCCGACCGG	CGGTTACTAT	ATTATGGTAA
•	551	AGAAAAGCGA	TGTGCGCGAA	CTCGATATGA	GCGTGGACGA	AGCGTTGAAA
40	601			GGTCATCCCT		
• •	651					GAACAACAAT
	701	AA				

### This corresponds to the amino acid sequence <SEQ ID 748; ORF98ng-1>:

```
45 MTEPAAEGGK AAKALKKYLI TGILVWLPIA VTVWVVSYIV SASDQLVNLL
51 PKQWRPQYVL GFNIPGLGVI VAIAVLFVTG LFAANVLGRQ ILAAWDSLLG
101 RIPVVKSIYS SVKKVSESLL SDSSRSFKTP VLVPFPQSGI WTIAFVSGQV
151 SNAVKAALPQ DGDYLSVYVP TTPNPTGGYY IMVKKSDVRE LDMSVDEALK
201 YVISLGMVIP DDLPVKTLAG PMPPEKAELP EOO*
```

#### ORF98ng-1 and ORF98-1 show 97.9% identity in 233 aa overlap:

50		10	20	30	40	50	60
	orf98-1.pep	MTEXAAEGGKAAKAL	KKYLITGI	LVWLPIAVTVW	VVSYIVSASI	OQLVNLLPKQW	IRPQYVL
			11111111	11111111111	111111111		
	orf98ng-1	MTEPAAEGGKAAKAL	KKYLITGI	LVWLPIAVTVW	VVSYIVSASI	<b>OQLVNLLPKQW</b>	RPQYVL
55		10	20	30	40	50	60
<i>JJ</i>		70	80	90	100	110	120
	orf98-1.pep	GFNIPGLGVIVAIAV					
		_	1111111		1111111111		111111

-419-

	orf98ng-1	GFNIPGLGVIVAIA	VLFVTGLFA	ANVLGRQILA/	AWDSLLGRIPV	VKSIYSSVKI	KVSESLL
		70	80	90	100	110	120
		130	140	150	160	170	180
5	orf98-1.pep	SDSSRSFKTPVLVP	FPQPGIWTI.	afvsgqvsna <sup>v</sup>	VKAALPKDGDY	(LSVYVPTTP)	NPTGGYY
•		111111111111111111111111111111111111111	111 11111		111111:111		11111
	orf98ng-1	SDSSRSFKTPVLVP	FPQSGIWTI:	AFVSGQVSNA <sup>1</sup>	VKAALPQDGDY	(LSVYVPTTP)	NPTGGYY
	•	130	140	150	160	170	180
10		190	200	210	220	230	
	orf98-1.pep	IMVKKSDVRELDMS	VDEALKYVI	SLGMVIPDDL:	PVKTLAGPMPS	SEKADLPEQQ	X
			111111111	1111111111			1
	orf98ng-1	IMVKKSDVRELDMS					X
		190	200	210	220	230	

Based on this analysis, including the fact that the putative transmembrane domains in the 15 gonococcal protein are identical to the sequences in the meningococcal protein, it is predicted that the proteins from N.meningitidis and N.gonorrhoeae, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 89

The following partial DNA sequence was identified in N.meningitidis <SEQ ID 749>: 20

```
ATGAAAACGG TAGTCTGGAT TGTCGTCCTG TTTGCCGCCG CCGTCGGACT
                    GGCGCTGGCT TCGGGCATTT ACACCGGCGA CGTGTATATC GTACTCGGAC
                 51
                    AGACCATGCT CAGAATCAAC CTGCACGCCT TTGTGTTAGG TTCGCTGATT
                101
                    GCCGTCGTGG TGTGGTATTT CTTGTTTAAA TTCATTATCG GgGTACTCA
                151
25
                    ATATCCCCGA AAAGATGCAG CGTTTCGGTT CGGCnCGTAA AGGCCkCAAG
                201
                    SSCGSGCTTG CCTTGAACAA GGCGGGTTTG GCGTATTTTG AAGGGCGTTT
                251
                    TGAAAAGGCG GAACTAGAAG CCTCACGCGT GTTGGTCAAC AAAGtAGGCC
                301
                    GaGAGACAAC CGGACTTTGG CATTGATGCT GrGCGCGCAC GCCGCCGGAC
                351
                    AGATGGAAAA CATCGASSTG CGCGACCGTT ATCTTGCGGA AATCGCCAAA
                401
30
                     CTGCCGGAAA AACAGCAGCT TTCCCGTTAT CTTTTGTTGG CGGAATCGGC
                451
                    GTTGAACCGG CGCGATTACG AAGCGGCGGA AGCCAATCTT CATGCGGCGG
                501
                     CGAAGATGAA TGCCAACCTT ACGCGCCTCG TGCGTCTGCA .ATTCGTTAC
                551
                     GCTTTCGACA GGGGCGACGC GTTGCAGGTT CTGGCAAAAA CCGAAAAACT
                601
                     TTCCAAGGCG GGCGCGTTGG GCAAATCGGA AATGGAACGG TATCAAAATT
                651
35
                     GGGCATATCC GTCGCCAGCT GGCGGATGCT GCCGATGCCG CCGCTTTGAA
                701
                     AACCTGCCTG AAGCGGATTC CCGACAGCCT CAAAAACGGG GAATTGAGCG
                751
                     TATCGGTTGC GGAAAAGTAC GAACGTTTGG GACTGTATGC CGATGCGGTC
                801
                851
                     AAATGGGTCA AACAGCATTA TCCGCASAAC CGCCGCCCCG AGCTTTTGGA
                     AGCCTTTGTC GAAAGCGTGC GCTTTTTGGG CGAGCGCGAA CAGCAGAAAG
                901
40
                951
                     CCATCGATTT TGCCGATGCT TGGCTGAAAG AACAGCCCGA TAACGCGCTT
                     CTGCTGATGT ATCTCGGTCG GCTCGCCTTC GGCCGCAAAC TTTGGGGCAA
               1001
               1051
                     GGCAAAAGGC TACCTTGAAG CGAGCATTGC ATTAAAGCCG AGTATTTCCG
                     CGCGTTTGGT TCTAACAAAG GTTTTCGACG AAATCGGAGA ACCGCAGAAG
               1101
                     GCGGAGGCGC AC...
```

45 This corresponds to the amino acid sequence <SEQ ID 750; ORF100>:

```
MKTVVWIVVL FAAAVGLALA SGIYTGDVYI VLGQTMLRIN LHAFVLGSLI
                    AVVVWYFLFK FIIGVLNIPE KMORFGSARK GXKXXLALNK AGLAYFEGRF
                    EKAELEASRV LVNKVGRDNR TLALMLXAHA AGQMENIXXR DRYLAEIAKL
               101
               151 PEKQQLSRYL LLAESALNRR DYEAAEANLH AAAKMNANLT RLVRLXIRYA
50
                    FDRGDALQVL AKTEKLSKAG ALGKSEMERY QNWAYRRQLA DAADAAALKT
                    CLKRIPDSLK NGELSVSVAE KYERLGLYAD AVKWVKQHYP XNRRPELLEA
               251
                    FVESVRFLGE REQOKAIDFA DAWLKEQPDN ALLLMYLGRL AFGRKLWGKA
               351 KGYLEASIAL KPSISARLVL TKVFDEIGEP QKAEAH...
```

Further work revealed the complete nucleotide sequence <SEQ ID 751>:

55	1	ATGAAAACGG	TAGTCTGGAT	TGTCGTCCTG	TTTGCCGCCG	CCGTCGGACT
	51	GGCGCTGGCT	TCGGGCATTT	ACACCGGCGA	CGTGTATATC	GTACTCGGAC
	101	AGACCATGCT	CAGAATCAAC	CTGCACGCCT	TTGTGTTAGG	TTCGCTGATT
	151	CCCCTCCTCC	<b>ጥርጥርርጥልጥጥ</b>	ርጥጥርጥጥጥ እ እ እ	ጥጥሮልጥጥልጥርር	CCCTACTCAA

	201 251 301	CCGCGCTTGC GAAAAGGCGG	CTTGAACAAG AACTAGAAGC	GCGGGTTTGG CTCACGCGTG	GGCGCGTAAA CGTATTTTGA TTGGTCAACA	AGGGCGTTTT
5	351 401 451 501	TGGAAAACAT CCGGAAAAAC	ACTTTGGCAT CGAGCTGCGC AGCAGCTTTC GATTACGAAG	GACCGTTATC CCGTTATCTT	TTGTTGGCGG	GCCGGACAGA CGCCAAACTG AATCGGCGTT GCGGCGGCGA
10	551 601 . 651	AGATGAATGC TTCGACAGGG CAAGGCGGGC	CAACCTTACG GCGACGCGTT GCGTTGGGCA	CGCCTCGTGC GCAGGTTCTG AATCGGAAAT	GTCTGCAACT GCAAAAACCG GGAACGGTAT	TCGTTACGCT AAAAACTTTC CAAAATTGGG
	701 751 801 851	TGCCTGAAGC GGTTGCGGAA	CCAGCTGGCG GGATTCCCGA AAGTACGAAC GCATTATCCG	CAGCCTCAAA GTTTGGGACT	AACGGGGAAT GTATGCCGAT	TTTGAAAACC TGAGCGTATC GCGGTCAAAT TTTGGAAGCC
15	901 951 1001	TTTGTCGAAA CGATTTTGCC TGATGTATCT	GCGTGCGCTT	TTTGGGCGAG TGAAAGAACA	CGCGAACAGC GCCCGATAAC GCAAACTTTG	AGAAAGCCAT
20	1051 1101 1151 1201	AAAGGCTACC TTTGGTTCTA AGGCGCAGCG GCAGCGTTAG	TTGAAGCGAG GCAAAGGTTT CAACTTGGTT AGCAGCATAG	TCGACGAAAT TTGGAAGCCG	AAGCCGAGTA CGGAGAACCG TCTCCGATGA	CAGAAGGCGG

This corresponds to the amino acid sequence <SEQ ID 752; ORF100-1>:

	1					LHAFVLGSLI
	51	AVVVWYFLFK	FIIGVLNIPE	KMQRFGSARK	GRKAALALNK	AGLAYFEGRE
25	. 101	EKAELEASRV	LVNKEAGDNR	TLALMLGAHA	AGQMENIELR	DRYLAEIAKL
	151	PEKQQLSRYL	LLAESALNRR	DYEAAEANLH	AAAKMNANLT	RLVRLQLRYA
	201	FDRGDALQVL	AKTEKLSKAG	ALGKSEMERY	ONWAYRROLA	DAADAAALKT
	251					HNRRPELLEA
	301					AYGRKLWGKA
30	351	KGYLEASIAL	KPSISARLVL	AKVFDEIGEP	OKAEAORNLV	LEAVSDDERH
	401	AALEQHS*			<b>L</b>	

Computer analysis of this amino acid sequence gave the following results:

# Homology with a predicted ORF from N.meningitidis (strain A)

ORF100 shows 93.5% identity over a 386aa overlap with an ORF (ORF100a) from strain A of N.

## 35 meningitidis:

		10	20	30	40	50	60
	orf100.pep						
	Offico.pep	MKTVVWIVVLFAAA					
		- 1111111111111			111111111		
	orf100a	MKTVVWIVVLFAAA	XGLALASGI	KTGDVYIVLGO	TMLRINLHAE	VLGSLTAVV	MYFT.FK
40		10	20	30	40	50	60
••			20	30	10	30	60
		70	80	90	100	110	100
	E100					110	120
	orf100.pep	FIIGVLNIPEKMOP	(FGSARKGXK	KXLALNKAGLA	YFEGRFEKAE	ELEASRVLVNI	KVGRDNR
		-		- 111111111			1:111
45	orf100a	FIIGVLNXPEKMQF	REGSARKGRK	AALALNKAGLA	YFEGREEKAR	TEASPULCN	CEACINE
		70	80	90	100	110	
		, ,	00	90	100	110	120
		130	140	150	160	170	180
	orf100.pep	TLALMLXAHAAGQN	MENIXXRDRY:	LAEIAKLPEKO	<b>QLSRYLLLAR</b>	SALNRRDYE	AAEANLH
50		111111 111111			1111111111	HILLIAN	111111
	orf100a	TLALMLGAHAAGON	KENTELDUDA		יווווווווווו מגדדזעמס זרע		
	0222000	120					
		130	140	150	160	170	180
		190	200	210	220	230	240
55	orf100.pep	AAAKMNANLTRLVF	RLXIRYAFDR	SDAT.OVT.AKTE	KLSKAGALGE	CSEMERYONW	A TODAY
·	orf100a		,, ,,,,,,,				
	Officua	AAAKMNANLTRLVI				(SEMERYQNW)	AYRRQLX
		190	200	210	220	230	240
60 .		250	260	270	280	290	300
	orf100.pep	DAADAAALKTCLKF					
	orrioo.beb			LOVOVALKILL	TOPINDAVKA	VVNQHIPXNRI	KPELLEA
			, , , , , , , , , , , ,				111111
	orf100a	DAADAAALKTCLKF	RIPDSLKNGE	LSVSVAEKYEF	RLGLYADAVKV	<b>VKQHYPHNR</b> I	RPELLEA
		250	260	270	280	290	300

		310 320 330 340 350 360
	orf100.pep	FVESVRFLGEREQQKAIDFADAWLKEQPDNALLLMYLGRLAFGRKLWGKAKGYLEASIAL
5	orf100a	FVESVRFLGERDQQKAIDFADAWLKEQPDNALLLXYLGRLAYGRKLWGKAKGYLEASIAL 310 320 330 340 350 360
		370 380
10	orf100.pep	KPSISARIVLTKVFDEIGEPOKAEAH
10	orf100a	KPSISARLVLAKVFDETGEPQKAEAQRNLVLASVAEENRPSAETHX
	m 1 / 141	370 380 390 400
		ORF100a nucleotide sequence <seq 753="" id=""> is:</seq>
15	1 ATC 51 GGC	GAAAACGG TAGTCTGGAT TGTCGTCCTG TTTGCCGCCG CNNTCGGGCT CATTGGCG TCGGGCATTN ACACCGGCGA CGTGTATATC GTACTCGGAC
13	101 AG	ACCATGCT CAGAATCAAC CTGCACGCCT TTGTGTTAGG TTCGCTGATT
		CGTCGTGG TGTGGTATTT CCTGTTCAAA TTCATCATCG GCGTACTCAA NCCCCGAA AAGATGCAGC GTTTCGGTTC GGCGCGTAAA GGCCGCAAGG
	251 CC	GCGCTTGC TTTGAACAAG GCGGGTTTGG CGTATTTTGA AGGGCGTTTT
20		AAAGGCGG AACTTGAAGC CTCGCGCGTA TTGGGAAACA AAGAGGCGGG ATAACCGG ACTTTGGCAT TGATGTTGGG CGCACATGCC GCCGGGCAGA
		GAAAACAT CGAGCTGCGC GACCGTTATC TTGCGGAAAT CGCCAAACTG
		GGAAAAGC AGCAGCTTTC CCGTTATCTT TTGTTGGCGG AATCGGCGTT
25		ACCGGCGC GATTACGAAG CGGCGGAAGC CAATCTTCAT GCGGCGGCGA ATGAATGC CAACCTTACG CGCCTCGTGC GTCTGCAACT TCGTTACGCT
25	601 TT	CGACAGGG GCGACGCGTT GCAGGTTCTG GCAAAAACCG AAAAANTTTC
		AGGCGGGC GCGTNGGGCA AATCGGAAAT GGAACGGTAT CAAAATTGGG TACCGCCG CCAGCTGNCG GATGCTGCCG ATGCCGCCGC TTTGAAAACC
	. •	CCTGAAGC GGATTCCCGA CAGCCTCAAA AACGGGGAAT TGAGCGTATC
30	<del>-</del>	TTGCGGAA AAGTACGAAC GTTTGGGACT GTATGCCGAT GCGGTCAAAT
	<del>-</del> · ·	GTCAAACA GCATTATCCG CACAACCGCC GACCCGAACT TTTGGAAGCN TGTCGAAA GCGTGCGCTT TTTGGGCGAA CGCGATCAGC AGAAAGCCAT
	. 951 CG	ATTTTGCC GATGCTTGGC TGAAAGAACA GCCCGATAAT GCGCTTCTGC
25	_	ANGTATCT CGGTCGGCTC GCCTACGGCC GCAAACTTTG GGGCAAGGCA AGGCTACC TTGAAGCGAG CATTGCATTA AAGCCGAGTA TTTCCGCGCG
35		TGGTTCTG GCAAAGGTTT TTGACGAAAC CGGAGAACCG CAGAAGGCGG
	1151 AG	GCGCAGCG CAACTTGGTT TTGGCAAGCG TTGCCGAGGA AAACCGNCCT
		CGCCGAAA CCCATTGA
	This encodes a prot	ein having amino acid sequence <seq 754="" id="">:</seq>
40		TVVWIVVL FAAAXGLALA SGIXTGDVYI VLGQTMLRIN LHAFVLGSLI
		VVWYFLFK FIIGVLNXPE KMORFGSARK GRKAALALNK AGLAYFEGRF AELEASRV LGNKEAGDNR TLALMLGAHA AGOMENIELR DRYLAEIAKL
		KQQLSRYL LLAESALNRR DYEAAEANLH AAAKMNANLT RLVRLQLRYA
4.5	201 FD	RGDALQVL AKTEKXSKAG AXGKSEMERY QNWAYRRQLX DAADAAALKT
45		KRIPDSLK NGELSVSVAE KYERLGLYAD AVKWVKQHYP HNRRPELLEA ESVRFLGE RDQOKAIDFA DAWLKEQPDN ALLLXYLGRL AYGRKLWGKA
	351 KG	YLEASIAL KPSISARLVL AKVFDETGEP QKAEAQRNLV LASVAEENRP
		ETH*
	ORF100a and ORF	100-1 show 95.1% identity in 406 aa overlap:
50		10 20 30 40 50 60
	orf100a.pep	MKTVVWIVVLFAAAXGLALASGIXTGDVYIVLGQTMLRINLHAFVLGSLIAVVVWYFLFK
	orf100-1	MKTVVWIVVLFAAAVGLALASGIYTGDVYIVLGQTMLRINLHAFVLGSLIAVVVWYFLFK
66		10 20 30 40 50 60
55		70 80 90 100 110 120
	orf100a.pep	FIIGVLNXPEKMQRFGSARKGRKAALALNKAGLAYFEGRFEKAELEASRVLGNKEAGDNR
	orf100-1	
60	011100-1	70 80 90 100 110 120
		130 140 150 160 170 180
	orf100a.pep	TLALMLGAHAAGQMENIELRDRYLAEIAKLPEKQQLSRYLLLAESALNRRDYEAAEANLH
65	orf100-1	
03	OLI100-1	I DVININGUIROZĀMINI TETUNKI PĀPTĀVPERĀŽĀTRUI BRITAROURIKUNI PĀĀPĀMIN

WO 99/24578 PCT/IB98/01665

-422-

		130	140	150	160	170	180
		190	200	210	220	230	240
_	orf100a.pep	AAAKMNANLTRLVRL	QLRYAFDRGD	alqvlakteky	SKAGAXGKSE	MERYQNWAYI	RRQLX
5			111111111	1111111	тин ши		1.111
	orf100-1	AAAKMNANLTRLVRL				MERYQNWAYI	RRQLA
		190	200	210	220	230	240
		252				6	
10	00	250	260	270	280	290	300
10	orf100a.pep	DAADAAALKTCLKRI	PDSLKNGELS	VSVAEKYERLO	<b>LYADAVKW</b> VK	QHYPHNRRPI	ELLEA
		<u> </u>		[			11111
	orf100-1	DAADAAALKTCLKRI					ELLEA
		250	260	270	280	290	300
1.5		212					
15		310	320	330	340	350	360
	orf100a.pep	FVESVRFLGERDOOK			LGRLAYGRKI	.WGKAKGYLE	ASIAL
						11111111	
	orf100-1	FVESVRFLGEREQQK					
20		310	320	330	340	350	360
20		0.70					
		370	380	390	400	•	
	orf100a.pep	KPSISARLVLAKVFD				X.	
	c1 00 1						
25	orf100-1	KPSISARLVLAKVFD				ISX	
25		370	380	390	400		

## Homology with a predicted ORF from N. gonorrhoeae

ORF100 shows 93.3% identity over a 386 aa overlap with a predicted ORF (ORF100ng) from

## N.gonorrhoeae:

30	orf100.pep	MKTVVWIVVLFAAAVGLALASGIYTGDVYIVLGQTMLRINLHAFVLG	SLIAVVVWYFLFK	60	
	orf100ng	MKTVVWIVVLFAAAVGLALASGIYTGDVYIVLGQTMLRINLHAFVLG	SLIAVVVWYFLFK	60	
35	orf100.pep	FIIGVLNIPEKMQRFGSARKGXKXXLALNKAGLAYFEGRFEKAELEA	SRVLVNKVGRDNR	120	
33	orf100ng		SRVLGNKEAGDNR	120	
	orf100.pep	TLALMLXAHAAGQMENIXXRDRYLAEIAKLPEKQQLSRYLLLAESAL	NRRDYEAAEANLH	180	
40	orf100ng	TLALMLGAHAAGQMENIELRDRYLAEIAKLPEKQQLSRYLLLAESAL		180	
	orf100.pep	AAAKMNANLTRLVRLXIRYAFDRGDALQVLAKTEKLSKAGALGKSEM	ERYQNWAYRRQLA	240	
45	orf100ng	AAAKMNANLTRLVRLQLRYAFDRGDALQVLAKTEKLSKAGALGKSEM		240	
43	orf100.pep	DAADAAALKTCLKRIPDSLKNGELSVSVAEKYERLGLYADAVKWVKQ	HYPXNRRPELLEA	300	
	orf100ng	DAADAAALKTCLKRIPDSLKNGELSVSVAEKYERLGLYADAVKWVKQ	HYPHNRRPELLEA	300	
50	orf100.pep	FVESVRFLGEREQQKAIDFADAWLKEQPDNALLLMYLGRLAFGRKLW	GKAKGYLEASIAL	360	
	orf100ng			360	
55	orf100.pep	KPSISARLVLTKVFDEIGEPQKAEAH	386		
55	orf100ng	:      ::      : KPSIPARLVLAKVFDETAQSQKAEAQRNLVLASVAGENRPSAETR	405		

# The complete length ORF100ng nucleotide sequence <SEQ ID 755> is:

	1	ATGAAAACGG	TAGTCTGGAT	TGTTGTCCTG	TTTGCCGCCG	CCGTCGGACT
	51	GGCGCTGGCT	TCGGGCATTT	ACACCGGCGA	CGTGTATATC	GTACTCGGAC
60	101	AGACCATGCT	CAGAATCAAC	CTGCACGCCT	TTGTGTTAGG	TTCGCTGATT
	151	GCCGTCGTGG	TGTGGTATTT	CCTGTTTAAA	TTCATCATCG	GCGTACTCAA
÷	201	TATCCCCGAA	AATATGCGGC	GTTCCGGTTC	GGCGCGGAAA	GGCCGCAAGG
	251	CCGCGCTTGC	CTTGAATAAG	GCGGGTTTGG	CGTATTTCGA	AGGGCGTTTT
	301	GAAAAGGCGG	AACTCGAAGC	CTCTCGAGTG	TTGGGCAACA	AAGAGGCCGG
65	351	AGACAACCGG	ACTTTGGCAT	TGATGCTGGG	CGCGCACGCG	GCAGGACAGA
	401				TTGCGGAAAT	

	501 AAAC 551 AGAT	AAAAAC AGCAGCTTTC ( CGGCGC GATTACGAAG ( GAATGC CAACCTTACG ( ATCGGG GCGATGCGTT (	CGGCGGAAGC CGCCTCGTGC	CAATCTTCAT GTCTGCAACT	GCGGCGGCGI TCGTTACGC	A C	
5	651 CAAG 701 CATA 751 TGCC	GCGGGG GCGATGCGTA GCGCCG GCGTTGGGCG GCGCAC GGATTGCCGA GCGGAA AAGTACGAAC	AATCGGAAAT GATGCTGCCG CAGCCTCAAA	GGAACGGTAT ATGCCGCCGC AACGGGGAAT	CAAAATTGGG TTTGAAAACG TGagcGTATG	3 C C	
10	851 GGGT 901 TTTG 951 CGAT 1001 TGAT	CAAACA GCATTATCCG TCGAAA GCGTGCGCTT TTTGCC GATTCTTGGC GTATCT CGGCCGGCTC	CACAACCGCC TTTGGGCGAG TGAAAGAACA GCCTACGGCC	GCCCGAGCT CGCGAACAGC GCCCGATAAC GCAAACTTTG	TTTGGAAGCCA' AGAAAGCCA' GCGCTTCTGG GGGTAAGGC	C T C A	
15	1101 TTTG 1151 AAGC	GCTACC TTGAAGCGAG GTGTTG GCAAAGGTTT ACAGCG CAACTTGGTT CCGAAA CCCGTTGA	TTGACGAAAC	CGCACAGTCG	CAAAAAGCC	G	
	This encodes a protein	n having amino acid	sequence <	SEQ ID 756	>:		
20	51 AVVV 101 EKAE 151 PEKÇ 201 FDRG	WWIVUL FAAAVGLALA WYFLFK FIIGULNIPE LEASRV LGNKEAGDNR QUSRYL LLAESALNRR DALQVL AKTEKLSKAG LIPDSLK NGELSVSVAE	NMRRSGSARK TLALMLGAHA DYEAAEANLH ALGKSEMERY	GRKAALALNK AGQMENIELR AAAKMNANLT ONWAYRROMA	AGLAYFEGR DRYLAEIAK RLVRLQLRY DAADAAALK	F L A T	
25	301 FVES	VRFLGE REQQKAIDFA LEASIAL KPSIPARLVL	DSWLKEQPDN	ALLLMYLGRL	AYGRKLWGK	A	
	ORF100ng and ORF	100-1 show 95.3% io	dentity in 40	2 aa overlap	<b>)</b> :		
30	orf100-1.pep	10 MKTVVWIVVLFAAAVGI	шинши	111111111111	1111111111	SLIAVVVWYFL	H
35	orf100-1.pep	70 FIIGVLNIPEKMQRFGS	11111111111	LNKAGLAYFEG	RFEKAELEAS	SRVLVNKEAGD	11
40	orf100-1.pep	70 130 TLALMLGAHAAGQMEN	80 140 IELRDRYLAEI	150 1 AKLPEKQQLSF	L60 RYLLLAESALI	170 1 NRRDYEAAEAN	.20 .80 NLH
45	orf100ng	TLALMLGAHAAGQMEN	IELRDRYLAEI 140	AKLPEKQQLSI	RYLLLAESALI	nrrdyeaaean	180 ITH
43	orf100-1.pep	190 AAAKMNANLTRLVRLQ             AAAKMNANLTRLVRLQ	THE HALL HAR LINE	QVLAKTEKLSI	KAGALGKSEM	ERYQNWAYRRQ	1:1
50	orf100ng	190	200		220	230 2	240
55	orf100-1.pep	250 DAADAAALKTCLKRIP             DAADAAALKTCLKRIP 250	111111111111	VAEKYERLGL 	YADAVKWVKQ             YADAVKWVKQ	HYPHNRRPELI             HYPHNRRPELI	111
60	orf100-1.pep	310  FVESVRFLGEREQQKA	320 AIDFADAWLKE	330 OPDNALLLMYL	340 GRLAYGRKLW	350 GKAKGYLEAS 	360 IAL 
65	orf100-1.pep	370 KPSISARLVLAKVFDE              KPSIPARLVLAKVFDET	:: 111111	RNLVLEAVSDD	:   :	Х	

-424-

370 380 390 400

Based on this analysis, including the presence of a putative leader sequence, a putative transmembrane domain, and a RGD motif, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 90

35

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 757>

```
1 ATGATGTTT CTTGGTTCAA GCTGTTTCAC TTGTTTTTTG TCATTTCGTG
10
                51 GTTTGCAGGG CTGTTTTACC TGCCGAGGAT TTTCGTCAAT ATGGCGATGA
               101 TTGATGTGCC GCGCGGCAAT CCCGAGTATG TGCGTCTGTC GGGCATGGCG
                   GTGCGGCTGT ACCGTTTTAT GTCGCCGTTG GGCTTCGGCG CGGTCGTGTT
               151
                    CGGCGCGCG ATACCGTTTG CCGCCGGCTG GTGGGGCAGC GGCTGGGTAC
               201
                    ACGTCAAACT GTGTTTGGGC TTGATGCTCT TGGCTTACCA GTTGTATTGC
               251
15
                    GGCGTGCTGC TGCGCCGTTT TCAGGATTAC AGCAATGCTT TTTCACACCG
               301
               351
                    CTGGTACCGC GTGTTCAACG AAATCCCCGT GCTGCTGATG GTTGCCGCGC
                    TGTATSTGGT CGTGTTCAAA CCGTTTTGA
```

This corresponds to the amino acid sequence <SEQ ID 758; ORF102>:

```
20 MMFSWFKLFH LFFVISWFAG LFYLPRIFVN MAMIDVPRGN PEYVRLSGMA
51 VRLYRFMSPL GFGAVVFGAA IPFAAGWWGS GWVHVKLCLG LMLLAYQLYC
101 GVLLRRFQDY SNAFSHRWYR VFNEIPVLLM VAALYXVVFK PF*
```

Further work revealed the complete nucleotide sequence <SEQ ID 759>:

```
ATGATGTTTT CTTGGTTCAA GCTGTTTCAC TTGTTTTTTG TCATTTCGTG
                51
                    GTTTGCAGGG CTGTTTTACC TGCCGAGGAT TTTCGTCAAT ATGGCGATGA
25
               101
                    TTGATGTGCC GCGCGGCAAT CCCGAGTATG TGCGTCTGTC GGGCATGGCG
                    GTGCGGCTGT ACCGTTTTAT GTCGCCGTTG GGCTTCGGCG CGGTCGTGTT
                151
                    CGGCGCGCG ATACCGTTTG CCGCCGGCTG GTGGGGCAGC GGCTGGGTAC
                201
                   ACGTCAAACT GTGTTTGGGC TTGATGCTCT TGGCTTACCA GTTGTATTGC
                251
                    GGCGTGCTGC TGCGCCGTTT TCAGGATTAC AGCAATGCTT TTTCACACCG
                301
30
                    CTGGTACCGC GTGTTCAACG AAATCCCCGT GCTGCTGATG GTTGCCGCGC
                351
                    TGTATCTGGT CGTGTTCAAA CCGTTTTGA
```

This corresponds to the amino acid sequence <SEQ ID 760; ORF102-1>:

```
1 MMFSWFKLFH LFFVISWFAG LFYLPRIFVN MAMIDVPRGN PEYVRLSGMA
51 VRLYRFMSPL GFGAVVFGAA IPFAAGWWGS GWVHVKLCLG LMLLAYQLYC
```

101 GVLLRRFQDY SNAFSHRWYR VFNEIPVLLM VAALYLVVFK PF\*

Computer analysis of this amino acid sequence gave the following results:

Homology with HP1484 hypothetical integral membrane protein of *H. pylori* (accession number AE000647) ORF102 and HP1484 show 33% aa identity in 143aa overlap:

```
FSWFKLFHLFFVISWFAGLFYLPRIFVNMAMIDVPRGNPEYVRLSGMAVRLYRFMSPLGF 62
          orf102 3
40
                      F W K FH+ VISW A LFYLPR+FV A
                                                             V++
          HP1484 8
                      FLWVKAFHVIAVISWMAALFYLPRLFVYHAENAHKKEFVGVVQIQEK--KLYSFIASPAM 65
                     GAVVFGAAIPFAAG---WWGSGWVHVKLCLGLMLLAYQLYCGVLLRRFQDYSNAFSHRWY 119
          orf102 63
                                      +
                                          GW+H KL L ++LLAY YC
                                                                +R +
45
                     GFTLITGILMLLIEPTLFKSGGWLHAKLALVVLLLAYHFYCKKCMRELEKDPTRRNARFY 125
          HP1484
          orf102 120 RVFNEIPXXXXXXXXXXXXFKPF 142
          HP1484 126 RVFNEAPTILMILIVILVVVKPF 148
```

## Homology with a predicted ORF from N. meningitidis (strain A)

ORF102 shows 99.3% identity over a 142aa overlap with an ORF (ORF102a) from strain A of N. meningitidis:

5		10	20	30	40	50	60
•	orf102.pep	MMFSWFKLFHLF	FVISWFAGLFYL	PRIFVNMAMI	DVPRGNPEYV	/RLSGMAVRLY	RFMSPL
	0111011	11111111111111	111111111111	111111111111	1111111111	1111111111	111111
	orf102a	MMFSWFKLFHLF	FVISWFAGLFYL	PRIFVNMAMI	DVPRGNPEY	/RLSGMAVRLY	RFMSPL
	011111	10	20	30	40	50	60
10						•	
10		70	80	. 90	100	110	120
	orf102.pep	GFGAVVFGAAIP	FAAGWWGSGWVH	VKLCLGLMLI	AYQLYCGVLI	LRRFQDYSNAF	SHRWYR
			111111111111	111111111	11111111		111111
	orf102a	GFGAVVFGAAIP	FAAGWWGSGWVH	VKLCLGLMLI			SHRWYR
15		70	80	90	100	110	120
		130	140				
	orf102.pep	VFNEIPVLLMVA	ALYXVVFKPFX				
		111111111111					
20	orf102a	VFNEIPVLLMVA	ALYLVVFKPFX				
		130	140				

The complete length ORF102a nucleotide sequence <SEQ ID 761> is:

```
25 101 TTGATGTCC GCGCGGCAAT CCCGAGGAT TTTCGTCAAT ATGGCGATGA
25 101 TTGATGTCC GCGCGGCAAT CCCGAGGAT TTTCGTCTC GGGCATGGCG
151 GTGCGGCTGT ACCGTTTTAT GTCGCCGTTG GGCTTCGGCG CGGTCGTGT
201 CGGCGCGGCG ATACCGTTTG CCGCCGCTG GTGGGGCAGC GGCTGGGTAC
251 ACGTCAAACT GTGTTTGGGC TTGATGCTCT TGGCTTACCA GTTGTATTGC
301 GGCGTGCTGC TGCGCCGTTT TCAGGATTAC AGCAATGCTT TTTCACACCG
301 TGTATCTGGT CGTGTTCAAAC CCGTTTTGA
```

This encodes a protein having amino acid sequence <SEQ ID 762>:

1 MMFSWFKLFH LFFVISWFAG LFYLPRIFVN MAMIDVPRGN PEYVRLSGMA
51 VRLYRFMSPL GFGAVVFGAA IPFAAGWWGS GWVHVKLCLG LMLLAYQLYC
35 101 GVLLRRFQDY SNAFSHRWYR VFNEIPVLLM VAALYLVVFK PF\*

ORF102a and ORF102-1 show complete identity in 142 aa overlap:

		10	20	30	40	50	60
	orf102a.pep	MMFSWFKLFHLFFVI	SWFAGLFYL	PRIFVNMAMI	DVPRGNPEYV	RLSGMAVRLY	RFMSPL
40	orf102-1	MMFSWFKLFHLFFVI	SWFAGLFYL	PRIFVNMAMI	DVPRGNPEYV		RFMSPL
		10	20	30	40	50	60
		70	80	90	100	110	120
45	orf102a.pep	GFGAVVFGAAIPFAA	AGWWGSGWVH	VKLCLGLMLL	AYQLYCGVLI	RRFQDYSNAF	SHRWYR
45	orf102-1	GFGAVVFGAAIPFA				RRFQDYSNAF	SHRWYR
		70	80	90	100	110	120
		130	140				
50	orf102a.pep	VFNEIPVLLMVAAL	/LVVFKPFX				
	orf102-1	VFNEIPVLLMVAAL	LVVFKPFX				

# 55 Homology with a predicted ORF from N.gonorrhoeae

ORF102 shows 97.9% identity over a 142 aa overlap with a predicted ORF (ORF102ng) from N. gonorrhoeae:

	orf102.pep orf102ng	MMFSWFKLFHLFFVISWFAGLFYLPRIFVNMAMIDVPRGNPEYVRLSGMAVRLYRFMSPL	60 60
5	orf102.pep		20
	orf102ng	GFGAVVFGAAIPFAAGRWGSGWVHVKLCLGLMLLAYQLYCGVLLRRFQDYSNAFSHRWYR 1	20
10	orf102.pep	VFNEIPVLLMVAALYXVVFKPF 142	
	The complete lengt	h ORF102ng nucleotide sequence <seq 763="" id=""> is:</seq>	
15	1 AT 51 GT 101 TT 151 GT 201 CG	GATGTTTT CTTGGTTCAA GCTGTTTCAC TTGTTTTTTG TCATTTCGTG TTGCAGGG CTGTTTTACC TGCCGAGGAT TTTCGTCAAT ATGGCGATGA GATGCGCC GCGCGGCAAT CCCGAGTATG TGCGCCTGTC GGGGATGGCG GCGGTTGT ACCGTTTTAT GTCGCCTTTG GGTTTCGGCG CGGTCGTGTT GCGCGGCG ATACCGTTTG CCGCcggccg GTGGGGCagc ggctggGTTC GTCAAACT GTGTTTGGGC TTGATGCTCT TGGCTTATCA GTTGTATTGC	
20	301 GG 351 CT	CGTGCTGC TGCGCCGTTT TCAGGATTAC AGCAATGCTT TTTCACACCG GGTACCGC GTGTTCAAcg aAATCCCCGT GCTGCTGATG GTTGCCGCGC TATCTGGT CGTGTTCAAA CCGTTTTGA	
		ein having amino acid sequence <seq 764="" id="">:</seq>	
25	51 VR	FSWFKLFH LFFVISWFAG LFYLPRIFVN MAMIDAPRGN PEYVRLSGMA LYRFMSPL GFGAVVFGAA IPFAAGRWGS GWVHVKLCLG LMLLAYQLYC LLRRFQDY SNAFSHRWYR VFNEIPVLLM VAALYLVVFK PF*	
	ORF102ng and OR	F102-1 show 98.6% identity in 142 aa overlap:	
30	orf102-1.pep	10 20 30 40 50 60  MMFSWFKLFHLFFVISWFAGLFYLPRIFVNMAMIDVPRGNPEYVRLSGMAVRLYRFMSPL	
35	orf102-1.pep	70 80 90 100 110 120 GFGAVVFGAAIPFAAGWWGSGWVHVKLCLGLMLLAYQLYCGVLLRRFQDYSNAFSHRWYR	
40	orf102-1.pep	130 140 VFNEIPVLLMVAALYLVVFKPFX	
	In addition, ORF10	2ng shows significant homology to a membrane protein from H.pylori:	
45	[Helicobacte Score = 79.	AE000647) conserved hypothetical integral membrane protein r pylori] Length = 148 2 bits (192), Expect = 1e-14 = 50/147 (34%), Positives = 68/147 (46%), Gaps = 13/147 (8%)	
50	Query: 3 F	SWFKLFHLFFVISWFAGLFYLPRIFVNMAMIDAPRGNPEYVRLSGMAVRLYRFMSPLGF 62 W K FH+ VISW A LFYLPR+FV A + V++ +LY F++ LWVKAFHVIAVISWMAALFYLPRLFVYHAENAHKKEFVGVVQIQEKKLYSFIASPAM 65	
55	Query: 63 G G Sbjct: 66 G	AVVFGAAIPFAAGRWGSGWVHVKLCLGLMLLAYQLYCGVLLRRFQDYSNAFS 115 + + F +G GW+H KL L ++LLAY YC +R + + FTLITGILMLLIEPTLFKSGGWLHAKLALVVLLLAYHFYCKKCMRELEKDPTRRN 121	
60		RWYRVFNEIPXXXXXXXXXXXFKPF 142 R+YRVFNE P KPF RFYRVFNEAPTILMILIVILVVVKPF 148	

Based on this analysis, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 91

The following partial DNA sequence was identified in N.meningitidis <SEQ ID 765>:

```
5
                   ATGGCAAAAA TGATGAAATG GGCGGCTGTT GCGGCGGTCG CGGCGGCAGC
                   GGTTTGGGGC GGATGGTCTT AACTGAAGCC CGAGCCGCAC GTGCTTGATA
                   TTACGGAAAC GGTCAGGCGC GGC // .....
               //.. ATTTCGTTTA CGATTTTGTC CGAACCGGAT ACGCCGATTA AGGCGAAGCT
               51 CGACAGCGTC GACCCCGGGC TGACCACGAT GTCGTCGGGC GGTTACAACA
10
              101 GCAGTACGGA TACGGCTTCC AATGCGGTCT ACTATTATGC CCGTTCGTTT
               151 GTGCCGAATC CGGACGGCAA ACTCGCCACG GGGATGACGA CGCAGAATAC
              201
                   GGTTGAAATC GACGGCGTGA AAAATGTGCT GATTATTCCG TCGCTGACCG
                   TGAAAAATCG CGGCGGCAAG GCGTTTGTGC GCGTGTTGGG TGCGGACGGC
              251
               301 AAGGCGCGG AACGCGAAAT CCGGACCGGT ATGAGAGACA GTATGAATAC
15
                   CGAAGTAAAA AGCGGGTTGA AAGAGGGGGA CAAAGTGGTC ATCTCCGAAA
               351
                   TAACCGCCGC CGAGCAACAG GAAAGCGGCG AACGCGCCCT AGGCGGCCCG
               401
               451 CCGCGCCGAT AA
     This corresponds to the amino acid sequence <SEQ ID 766; ORF85>:
                   MAKMMKWAAV AAVAAAAVWG GWS.LKPEPH VLDITETVRR G......
20
                51
                    101
                   151
                   201
                   ...... I SFTILSEPDT
                   PIKAKLDSVD PGLTTMSSGG YNSSTDTASN AVYYYARSFV PNPDGKLATG
                   MTTQNTVEID GVKNVLIIPS LTVKNRGGKA FVRVLGADGK AAEREIRTGM
25
               301
               351 RDSMNTEVKS GLKEGDKVVI SEITAAEQQE SGERALGGPP RR*
     Further work revealed the further partial nucleotide sequence <SEQ ID 767>:
                   ..GTATCGGTCG GCGCGCAGGC ATCGGGGCAG ATTAAGATAC TTTATGTCAA
                51
                     ACTCGGGCAA CAGGTTAAAA AGGGCGATTT GATTGCGGAA ATCAATTCGA
30
               101
                     CCTCGCAGAC CAATACGCTC AATACGGAAA AATCCAAGTT GGAAACGTAT
               151
                     CAGGCGAAGC TGGTGTCGGC ACAGATTGCA TTGGGCAGCG CGGAGAAGAA
                     ATATAAGCGT CAGGCGGCGT TATGGAAGGA AAACGCGACT TCCAAAGAGG
               201
               251
                     ATTTGGAAAG CGCGCAGGAT GCGTTTGCCG CCGCCAAAGC CAATGTTGCC
               301
                     GAGCTGAAGG CTTTAATCAG ACAGAGCAAA ATTTCCATCA ATACCGCCGA
35
               351
                     GTCGGAATTG GGCTACACGC GCATTACCGC AACGATGGAC GGCACGGTGG
                     TGGCGATTCT CGTGGAAGAG GGGCAGACTG TGAACGCGGC GCAGTCTACG
               401
               451
                     CCGACGATTG TCCAATTGGC GAATCTGGAT ATGATGTTGA ACAAAATGCA
                     GATTGCCGAG GGCGATATTA CCAAGGTGAA GGCGGGGCAG GATATTTCGT
               501
               551
                     TTACGATTTT GTCCGAACCG GATACGCCGA TTAAGGCGAA GCTCGACAGC
                     GTCGACCCCG GGCTGACCAC GATGTCGTCG GGCGGTTACA ACAGCAGTAC
40
               601
               651
                     GGATACGGCT TCCAATGCGG TCTACTATTA TGCCCGTTCG TTTGTGCCGA
               701
                     ATCCGGACGG CAAACTCGCC ACGGGGATGA CGACGCAGAA TACGGTTGAA
               751
                     ATCGACGGCG TGAAAAATGT GCTGATTATT CCGTCGCTGA CCGTGAAAAA
                     TCGCGGCGGC AAGGCGTTTG TGCGCGTGTT GGGTGCGGAC GGCAAGGCGG
               801
45
               851
                     CGGAACGCGA AATCCGGACC GGTATGAGAG ACAGTATGAA TACCGAAGTA
               901
                     AAAAGCGGGT TGAAAGAGGG GGACAAAGTG GTCATCTCCG AAATAACCGC
               951
                      CGCCGAGCAA CAGGAAAGCG GCGAACGCGC CCTAGGCGGC CCGCCGCGCC
              1001
                      GATAA
      This corresponds to the amino acid sequence <SEQ ID 768; ORF85-1>:
```

50	1	VSVGAQASGQ	IKILYVKLGQ	QVKKGDLIAE	INSTSQTNTL	NTEKSKLETY
	51	QAKLVSAQIA	LGSAEKKYKR	.QAALWKENAT	SKEDLESAQD	AFAAAKANVA
	101	ELKALIRQSK	ISINTAESEL	GYTRITATMD	GTVVAILVEE	GQTVNAAQST
	151	PTIVQLANLD	MMLNKMQIAE	GDITKVKAGQ	DISFTILSEP	DTPIKAKLDS
	201	VDPGLTTMSS	GGYNSSTDTA	SNAVYYYARS	FVPNPDGKLA	TGMTTQNTVE
55	251	IDGVKNVLII	PSLTVKNRGG	KAFVRVLGAD	GKAAEREIRT	GMRDSMNTEV
	301	KSGLKEGDKV	VISEITAAEQ	QESGERALGG	PPRR*	

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from N. meningitidis (strain A)

ORF85 shows 87.8% identity over a 41aa overlap and 99.3% identity over a 153aa overlap with an ORF (ORF85a) from strain A of N. meningitidis:

```
5
                      MAKMMKWAAVAAVAAAVWGGWS-LKPEPHVLDITETVRRG
          orf85.pep
                      orf85a
                      MAKMMKWAAVAAAAAVWGGWSYLKPEPQAAYITETVRRGDISRTVSATGEISPSNLVS
                              10
                                        20
                                                 30
                                                           40
                                                                     50
10
                                                             90
                                                                      100
                                                   80
          orf85.pep
                                                   . ISFTILSEPDTPIKAKLDSVDPGLTTMSSG
                                                    111111111111111111111111111111111111
          orf85a
                      TIVQLANLDMMLNKMQIAEGDITKVKAGQDISFTILSEPDTPIKAKLDSVDPGLTTMSSG
                    210
                              220
                                        230
                                                  240
                                                           250
                                                                     260
15
                               120
                      110
                                         130
                                                  140
                                                            150
                      GYNSSTDTASNAVYYYARSFVPNPDGKLATGMTTQNTVEIDGVKNVLIIPSLTVKNRGGK
          orf85.pep
                       GYNSSTDTASNAVYYYARSFVPNPDGKLATGMTTQNTVEIDGVKNVLIIPSLTVKNRGGR
          orf85a
20
                     270
                              280
                                        290
                                                 300
                                                           310
                      170
                               180
                                         190
                                                  200
                                                            210
                                                                      220
          orf85.pep
                       AFVRVLGADGKAAEREIRTGMRDSMNTEVKSGLKEGDKVVISEITAAEQQESGERALGGP
                       25
          orf85a
                       AFVRVLGADGKAAEREIRTGMRDSMNTEVKSGLKEGDKVVISEITAAEQQESGERALGGP
                     330
                              340
                                        350
                                                  360
                                                           370
                                                                     380
                      230
                       PRRX
          orf85.pep
30
                       1111
          orf85a
                       PRRX
                     390
     The complete length ORF85a nucleotide sequence <SEO ID 769> is:
                    ATGGCAAAAA TGATGAAATG GGCGGCTGTT GCGGCGGTCG CGGCGGCAGC
35
                51
                    GGTTTGGGGC GGATGGTCTT ATCTGAAGCC CGAGCCGCAG GCTGCTTATA
               101
                    TTACGGAAAC GGTCAGGCGC GGCGACATCA GCCGGACGGT TTCTGCAACA
               151
                    GGGGAGATTT CGCCGTCCAA CCTGGTATCG GTCGGCGCGC AGGCATCGGG
               201
                    GCAGATTAAG AAACTTTATG TCAAACTCGG GCAACAGGTT AAAAAGGGCG
               251
                    ATTTGATTGC GGAAATCAAT TCGACCTCGC AGACCAATAC GCTCAATACG
40
               301
                    GAAAAATCCA AATTGGAAAC GTATCAGGCG AAGCTGGTGT CGGCACAGAT
               351
                    TGCATTGGGC AGCGCGGAGA AGAAATATAA GCGTCAGGCG GCGTTGTGGA
               401
                    AGGATGATGC GACCGCTAAA GAAGATTTGG AAAGCGCACA GGATGCGCTT
               451
                    GCCGCCGCCA AAGCCAATGT TGCCGAGCTG AAGGCTCTAA TCAGACAGAG
               501
                    CAAAATTTCC ATCAATACCG CCGAGTCGGA ATTGGGCTAC ACGCGCATTA
45
               551
                    CCGCAACGAT GGACGCCACG GTGGTGGCGA TTCTCGTGGA AGAGGGGCAG
               601
                    ACTGTGAACG CGGCGCAGTC TACGCCGACG ATTGTCCAAT TGGCGAATCT
               651
                    GGATATGATG TTGAACAAAA TGCAGATTGC CGAGGGCGAT ATTACCAAGG
               701
                    TGAAGGCGGG GCAGGATATT TCGTTTACGA TTTTGTCCGA ACCGGATACG
               751
                    CCGATTAAGG CGAAGCTCGA CAGCGTCGAC CCCGGGCTGA CCACGATGTC
50
                    GTCGGGCGGC TACAACAGCA GTACGGATAC GGCTTCCAAT GCGGTCTACT
               801
               851
                    ATTATGCCCG TTCGTTTGTG CCGAATCCGG ACGGCAAACT CGCCACGGGG
                    ATGACGACGC AGAATACGGT TGAAATCGAC GGTGTGAAAA ATGTGCTGAT
               901
               951
                    TATTCCGTCG CTGACCGTGA AAAATCGCGG CGGCAGGGCG TTTGTGCGCG
                    TGTTGGGTGC AGACGGCAAG GCGGCGGAAC GCGAAATCCG GACCGGTATG
              1001
55
              1051
                    AGAGACAGTA TGAATACCGA AGTAAAAAGC GGGTTGAAAG AGGGGGACAA
                    AGTGGTCATC TCCGAAATAA CCGCCGCCGA GCAGCAGGAA AGCGGCGAAC
              1101
                    GCGCCCTAGG CGGCCGCCG CGCCGATAA
              1151
      This encodes a protein having amino acid sequence <SEQ ID 770>:
                 1
                    MAKMMKWAAV AAVAAAAVWG GWSYLKPEPQ AAYITETVRR GDISRTVSAT
60
                51
                    GEISPSNLVS VGAQASGQIK KLYVKLGQQV KKGDLIAEIN STSQTNTLNT
               101
                    EKSKLETYQA KLVSAQIALG SAEKKYKRQA ALWKDDATAK EDLESAQDAL
```

AAAKANVAEL KALIRQSKIS INTAESELGY TRITATMDGT VVAILVEEGQ

TVNAAQSTPT IVQLANLDMM LNKMQIAEGD ITKVKAGQDI SFTILSEPDT PIKAKLDSVD PGLTTMSSGG YNSSTDTASN AVYYYARSFV PNPDGKLATG

MTTONTVEID GVKNVLIIPS LTVKNRGGRA FVRVLGADGK AAEREIRTGM

151 201

251

65

351 RDSMNTEVKS GLKEGDKVVI SEITAAEQQE SGERALGGPP RR\* ORF85a and ORF85-1 show 98.2% identity in 334 aa overlap:

		30	40	50	60	70	80	
	orf85a.pep	PQAAYIT	ETVRRGDI	SRTVSATGEI	SPSNLVSVGA	QASGQIKKLY	VKLGQQVKKG	DLIAE
5							VKLGQQVKKG	
	orf85-1				VSVGA	DASGOIRIDI 10	20	30
		90	100	110	120	130	140	TOROD
10	orf85a.pep	INSTSQT	NTLNTEKS	KLETYQAKLV	SAQIALGSAE	KKYKRQAALW	KUDATAKEDI	IIIIIII IIIIIII
	orf85-1	TNSTSOT	NTLNTEKS	KLETYOAKLV	SAQIALGSAE	KKYKROAALW	KENATSKEDI	ESAQD
	OLIOS I	11.01.02.	40	50	60	70	80	90
							'000	
15		150	160	170	180 AESELGYTRI	190	200 TTVEECOWN	<b>ጥ</b> ዖ∩ፈፈኒ
	orf85a.pep	ALAAAKA I · I I I I I	UNVAEDKAL	TKÖSVISINI		IIIIIIIIII	THAFFGGIAL	
	orf85-1	AFAAAKA	NVAELKAI	IRQSKISIN	AESELGYTRI	TATMDGTVVA	LILVEEGQTV	TROAAN
	01100 -		100	110	120	130	140	150
20			•••	000	040	250	260	
	£0.5 a mam	210	220	230 MOTAECDITE	240 KVKAGQDISFT			LTTMSS
	orf85a.pep	1111111						
	orf85-1	PTIVQLA	ANLDMMLN	MQIAEGDIT!	KVKAGQDISFT	ILSEPDTPIE	KAKLDSVDPG	LTTMSS
25			160	170	180	190	200	210
		270	280	290	300	310	320	
	orf85a.pep				PDGKLATGMTI			VKNRGG
	Ollosa.pop	111111	11111111	111111111	1111111111	111111111	111111111	11111
30	orf85-1	GGYNSS			PDGKLATGMTT	QNTVEIDGVI	KNVLIIPSLT	VKNRGG 270
			220	230	240	250	260	270
		330	340	350	360	370	380	
	orf85a.pep	RAFVRV	LGADGKAA	EREIRTGMRD	SMNTEVKSGL	(EGDKVVISE	ITAAEQQESG	ERALGG
35		:11111	1111111			TO PROPERTY OF THE		FDALCC
	orf85-1	KAFVRV.	LGADGKAA 280	EREIRTGMRD	SMNTEVKSGLE 300	310	11AAEQQESG 320	330
			200	230	300	310	525	
		390						
40	orf85a.pep	PPRRX						
	orf85-1	  PPRRX						
	¥=							
	Figure 10D chows	nlote of	hvdronhi	licity antio	enic index	and AMPF	II regions f	or ORF8

Figure 19D shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF85a..

# Homology with a predicted ORF from N.gonorrhoeae

ORF85 shows a high degree of identity with a predicted ORF (ORF85ng) from N.gonorrhoeae:

	ORF85	1	MAKMMKWAAVAAAAVWGGWS.LKPEPHVLDITETVRRG	40
	ORF85ng	1	MAKMMKWAAVAAVAAAAVWGGWSYLKPEPQAAYITEAVRRGDISRTVSAT	50
50				
30	ORF85		ISFTILSEPDT	250
	ORF85ng	201	TVNAAQSTPTIVQLANLDMMLNKMQIAEGDITKVKAGQDISFTILSEPDT	250
55	ORF85	251	PIKAKLDSVDPGLTTMSSGGYNSSTDTASNAVYYYARSFVPNPDGKLATG	300
	ORF85ng	251	PIKAKLDSVDPGLTTMSSGGYNSSTDTASNAVYYYARSFVPNPDGKLATG	300
60	ORF85	301	MTTONTVEIDGVKNVLIIPSLTVKNRGGKAFVRVLGADGKAAEREIRTGM	350
00	ORF85ng	301	MTTQNTVEIDGVKNVLLIPSLTVKNRGGKAFVRVLGADGKAVEREIRTGM	350
	ORF85	152	RDSMNTEVKSGLKEGDKVVISEITAAEQQESGERALGGPPRR 393	
65	ORF85ng	351	KDSMNTEVKSGLKEGDKVVISEITAAEQQESGERALGGPPRR 393	

# The complete length ORF85ng nucleotide sequence <SEQ ID 771> is:

	1 51				GCGGCGGTCG CGAACCGCAG		
	101				GCCGGACGGT		
5	. 151	GgcgAGATTT	CGCCGTCCAA	CCTGGTATCG	GTCGGCGCGC	AGGCTTCGG	G .
	201				GCAACAGGTC		
	251				AGACCAACAC		
	. 301 351				AAGCTGGTGT GCGTCAGGCG		
10	401				AAAGCGCGCA		
10	451				AAGGCTTTAA		
	501				TTTGGGCTAC		
	551	CCGCGACGAT	<b>GGACGGCACG</b>	GTGGTGGCGA	TTCCCGTGGA	AGAGGGGCA	AG
	601				ATTGTCCAAT		
15	651				CGAGGGCGAT		
	701				TTTTGTCCGA		
	751 801				CCCGGGCTGA GGCTTCCAAT		
	851	ATTATECCCE	TACAACAGCA	CCAPATCCC	ACGGCAAACT	CCCCACCC	1°1 2¢
20	901	ATGACGACGC	AGAATACGGT	TGAAATCGAC	GGTGTGAAAA	ATGTGTTGC	-πρ -πρ
20	951	TATTCCGTCG	CTGACCGTGA	AAAATCGCGG	CGGCAAGGCG	TTCGTACGC	CG
	1001	TGTTGGGTGC	GGACGGCAAG	GCAGTGGAAC	GCGAAATCCG	GACCGGTAT	ľG
	1051	AAAGACAGTA	TGAATACCGA	AGTGAAAAGC	GGGTTGAAAG	AGGGGGACA	AA .
0.5	1101				GCAGCAGGAA	AGCGGCGAF	AC .
25	1151	GCGCCCTAGG	CGGCCCGCCG	CGCCGATAA		•	
	This encodes a	protein havin	g amino acio	d sequence <	SEQ ID 772	>:	
	1	MAKMMKWAAV	AAVAAAAVWG	GWSYLKPEPO	AAYITEAVRR	GDISRTVSA	ΑT
	51				KKGDLIAEIN		
••	101				ALWKDDATSK		
30	151				TRITATMOGT		
•	201 251				ITKVKAGQDI		
	301				AVYYYARSFV FVRVLGADGK		
	351	KDSMNTEVKS					314
			GTUTEGDVAAT	DETINATION	SULKALGGER	KK"	
35	ORF85ng and C			,		KK*	
35		ORF85-1 sho	w 96.1% ide	ntity in 334	aa overlap:		
35	ORF85ng and O	ORF85-1 sho	w 96.1% ide	ntity in 334	aa overlap:	70	80
35		ORF85-1 sho	w 96.1% ide	ntity in 334	aa overlap: 60 vsvgaqasgq1	70 KKLYVKLGQQ	QVKKGDLIAE
35	ORF85ng and C	ORF85-1 sho	w 96.1% ide	ntity in 334	aa overlap:  60 VSVGAQASGQI	70 KKLYVKLGQ(	QVKKGDLIAE
35 40	ORF85ng and O	ORF85-1 sho	w 96.1% ide	ntity in 334	aa overlap:  60 VSVGAQASGQI	70 KKLYVKLGQQ           KILYVKLGQ	QVKKGDLIAE             QQVKKGDLIAE
	ORF85ng and C	ORF85-1 sho	w 96.1% ide 40 tvrrgdisrtv	ntity in 334 50 SATGEISPSNI	aa overlap:  60 VSVGAQASGQI            VSVGAQASGQ	70 KKLYVKLGQQ           KILYVKLGQ	QVKKGDLIAE             QQVKKGDLIAE
	ORF85ng and O	ORF85-1 shows the state of the	w 96.1% ide 40 TVRRGDISRTV 100	ontity in 334 50 SATGEISPSNI	aa overlap:  60 .vsvgAQASGQI             vsvgAQASGQ 10  120 1	70 KKLYVKLGQ(            KILYVKLG( 20	QVKKGDLIAE            QQVKKGDLIAE 0 30
	ORF85ng and C	ORF85-1 shows the state of the	w 96.1% ide 40 tvrrgdisrtv 100 ntidmekskle	ontity in 334  50 SATGEISPSNI  110 TYQAKLVSAQI	aa overlap:  60 .VSVGAQASGQI             VSVGAQASGQI 10 120 120 1ALGSAEKKYKE	70 KKLYVKLGQ(           IKILYVKLG( 20 30 QAALWKDDA	QVKKGDLIAE             QQVKKGDLIAE 0 30  140 TSKEDLESAQD
40	ORF85ng and O	ORF85-1 shows 30 PQAAYITE 90 INSTTOT	w 96.1% ide  40  TVRRGDISRTV  100  NTIDMEKSKLE	ontity in 334  50 SATGEISPSNI  110 TYQAKLVSAQI	aa overlap:  60 VSVGAQASGQI VSVGAQASGQI 10 120 1ALGSAEKKYKR	70 KKLYVKLGQ(                IKILYVKLG( 20 30 QAALWKDDA	QVKKGDLIAE              QQVKKGDLIAE 0 30  140 TSKEDLESAQD
	ORF85ng and O	ORF85-1 shows 30 PQAAYITE 90 INSTTOT	w 96.1% ide  40  TVRRGDISRTV  100  NTIDMEKSKLE      : :              NTLNTEKSKLE	ntity in 334  50 SATGEISPSNI  110 TYQAKLVSAQI	aa overlap:  60 VSVGAQASGQI VSVGAQASGQI 10 120 1ALGSAEKKYKR	70 KKLYVKLGQ(                IKILYVKLG( 20 30 QAALWKDDA:	QVKKGDLIAE             QQVKKGDLIAE 0 30  140 TSKEDLESAQD
40	ORF85ng and O	ORF85-1 shows 30 PQAAYITE 90 INSTTOT	w 96.1% ide  40  TVRRGDISRTV  100  NTIDMEKSKLE	ntity in 334  50 SATGEISPSNI  110 TYQAKLVSAQI	aa overlap:  60 VSVGAQASGQI VSVGAQASGQI 10 120 1ALGSAEKKYKR	70 KKLYVKLGQ(                IKILYVKLG( 20 30 QAALWKDDA:	QVKKGDLIAE             QQVKKGDLIAE 0 30  140 TSKEDLESAQD
40	ORF85ng and O	ORF85-1 shows 30 PQAAYITE 90 INSTTOT	w 96.1% ide  40  TVRRGDISRTV  100  NTIDMEKSKLE      : :              NTLNTEKSKLE	ntity in 334  50 SATGEISPSNI  110 TYQAKLVSAQI	aa overlap:  60 VSVGAQASGQI VSVGAQASGQI 10 120 120 1ALGSAEKKYKR IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	70  KKLYVKLGQ(                IKILYVKLG( 20  30 QAALWKDDA'	QVKKGDLIAE             QQVKKGDLIAE 0 30  140 TSKEDLESAQD
40	ORF85ng and O	ORF85-1 shows 30 PQAAYITE 90 INSTTOT INSTSOT 150 ALAAAKA	w 96.1% ide  40  TVRRGDISRTV  100  NTIDMEKSKLE      :	ntity in 334  50 SATGEISPSNI  110 TYQAKLVSAQI 1111111111 TYQAKLVSAQI 50 6 170 SKISINTAESI	aa overlap:  60 VSVGAQASGQI            VSVGAQASGQI 10  120 1ALGSAEKKYKR	70 KKLYVKLGQ(                IKILYVKLG( 20 30 QAALWKDDA'             :     QAALWKENA' 80	QVKKGDLIAE             QQVKKGDLIAE 0 30  140 TSKEDLESAQD             TSKEDLESAQD 0 90  200 EGQTVNAAQST
40	ORF85ng and O orf85ng orf85-1 orf85ng orf85-1 orf85ng	ORF85-1 shows 30 PQAAYITE  90 INSTTOT	w 96.1% ide  40  TVRRGDISRTV  100  NTIDMEKSKLE  11::        NTLNTEKSKLE  40  160  NVAELKALIRO	ntity in 334  50 SATGEISPSNI  110 TYQAKLVSAQI 111111111111111111111111111111111111	aa overlap:  60 VSVGAQASGQI             VSVGAQASGQI 10  120 120 1ALGSAEKKYKR            ALGSAEKKYKR 60 70  180 1DLGYTRITATME	70 KKLYVKLGQQ                   IKILYVKLGQ 20 30 QAALWKDDA:                   QAALWKENA: 80 90 GTVVAIPVE	QVKKGDLIAE
40	ORF85ng and O orf85ng orf85-1 orf85ng orf85-1	ORF85-1 shows 30 PQAAYITE  90 INSTTOT	w 96.1% ide  40  TVRRGDISRTV  100  NTIDMEKSKLE   ::       NTLNTEKSKLE 40  160  NVAELKALIRO	ntity in 334  50 SATGEISPSNI  110 TYQAKLVSAQI            TYQAKLVSAQI 50  170 SKISINTAESE	aa overlap:  60 VSVGAQASGQI             VSVGAQASGQI 10  120 120 1ALGSAEKKYKR            ALGSAEKKYKR 60 70  180 1 OLGYTRITATME	70  KKLYVKLGQQ                  IKILYVKLGQ  20  30  QAALWKDDA'                  QAALWKENA'  80  90  GTVVAIPVE	QVKKGDLIAE             QQVKKGDLIAE 0 30  140 TSKEDLESAQD             TSKEDLESAQD 0 90  200 EGQTVNAAQST
40	ORF85ng and O orf85ng orf85-1 orf85ng orf85-1 orf85ng	ORF85-1 shows 30 PQAAYITE  90 INSTTOT	w 96.1% ide  40  TVRRGDISRTV  100  NTIDMEKSKLE   ::       NTLNTEKSKLE 40  160  NVAELKALIRO	ntity in 334  50 SATGEISPSNI  110 TYQAKLVSAQI 111111111111111111111111111111111111	aa overlap:  60 VSVGAQASGQI             VSVGAQASGQI 10  120 120 1ALGSAEKKYKR            ALGSAEKKYKR 60 70  180 1 OLGYTRITATME	70  KKLYVKLGQQ                  IKILYVKLGQ  20  30  QAALWKDDA'                  QAALWKENA'  80  90  GTVVAIPVE	QVKKGDLIAE             QQVKKGDLIAE 0 30  140 TSKEDLESAQD             TSKEDLESAQD 0 90  200 EGQTVNAAQST
40	ORF85ng and O orf85ng orf85-1 orf85ng orf85-1 orf85ng	30 PQAAYITE  90 INSTTOT IIII:II INSTSOT  150 ALAAAKA I:IIIII AFAAAKA	w 96.1% ide  40  TVRRGDISRTV  100  NTIDMEKSKLE  !!:: !!!!!  NTLNTEKSKLE  40  160  NVAELKALIRO  III!!!!!!!  NVAELKALIRO  100  1	ontity in 334  50 SATGEISPSNI  110 TYQAKLVSAQI	aa overlap:  60 .vsvGAQASGQI             vsvGAQASGQI 10  120 120 1ALGSAEKKYKR             ALGSAEKKYKR 60 70  180 1DLGYTRITATME	70  KKLYVKLGQ(                    IKILYVKLG( 20  30  QAALWKDDA:                  QAALWKENA*   80  GTVVAIPVE                 GTVVAILVE    14	QVKKGDLIAE             QQVKKGDLIAE 0 30  140 TSKEDLESAQD             TSKEDLESAQD 0 90  200 EGQTVNAAQST            EGQTVNAAQST
40 45 50	ORF85ng and O orf85ng orf85-1 orf85ng orf85-1 orf85ng orf85-1	30 PQAAYITE  90 INSTTOT IIII:II INSTSOT  150 ALAAAKA I:IIIII AFAAAKA	w 96.1% ide  40  TVRRGDISRTV  100  NTIDMEKSKLE  !!:: !!!!!  NTLNTEKSKLE  40  160  NVAELKALIRO  !!!!!!!!!!  NVAELKALIRO  100  1	ontity in 334  50 SATGEISPSNI  110 TYQAKLVSAQI             TYQAKLVSAQI 50 25              SKISINTAESE             SKISINTAESE 10 12	aa overlap:  60 .vsvgaQasgQI            vsvgaQasgQI 10  120 120 1ALGSAEKKYKR              ALGSAEKKYKR              ALGSAEKTYKR 10 180 10 10 110 110 110 110 110 110 11	70  KKLYVKLGQ(                    IKILYVKLG( 20  30  QAALWKDDA:                    QAALWKENA:   80  90  GTVVAIPVE:                GTVVAILVE:   14	QVKKGDLIAE             QQVKKGDLIAE 0 30  140 TSKEDLESAQD             TSKEDLESAQD 0 90  200 EGQTVNAAQST            EGQTVNAAQST 0 150
40	ORF85ng and O orf85ng orf85-1 orf85ng orf85-1 orf85ng	30 PQAAYITE  90 INSTTOT III:II INSTSOT  150 ALAAAKA I:IIIII AFAAAKA  210 PTIVQLA	W 96.1% ide  40  TVRRGDISRTV  100  NTIDMEKSKLE  !!::!!!!!  NTLNTEKSKLE  40  160  NVAELKALIRO                NVAELKALIRO  100  220  NLDMMLNKMQI	ntity in 334  50 SATGEISPSNI  110 TYQAKLVSAQI             TYQAKLVSAQI 50 250 270 SKISINTAESE            230 AEGDITKVKAG	aa overlap:  60 .vsvgaQasgQI            vsvgaQasgQI 10  120 120 1ALGSAEKKYKR              ALGSAEKKYKR              ALGSAEKTYKR 10 180 10 10 110 110 110 110 110 110 11	70  KKLYVKLGQ(                    IKILYVKLG( 20  30  QAALWKDDA:                    QAALWKENA:   80  90  GTVVAIPVE:                GTVVAILVE:   14	QVKKGDLIAE
40 45 50	ORF85ng and O orf85ng orf85-1 orf85ng orf85-1 orf85ng orf85-1	90 INSTTOT INSTSOT  150 ALAAAKA I:     AFAAAKA  210 PTIVQLA	W 96.1% ide  40  TVRRGDISRTV  100  NTIDMEKSKLE    ::       NTLNTEKSKLE  40  160  NVAELKALIRO  100  1  220  NLDMMLNKMOI	ntity in 334  50 SATGEISPSNI  110 TYQAKLVSAQI                         TYQAKLVSAQI 50 270 SKISINTAESE                       SKISINTAESE                     230 AEGDITKVKAG	aa overlap:  60 VSVGAQASGQI             VSVGAQASGQI 10  120 1 ALGSAEKKYKR             ALGSAEKKYKR 60 70  180 1 ALGYTRITATME            CLGYTRITATME 20 130  240 2 EQDISFTILSER	70  KKLYVKLGQ(                 IKILYVKLG( 20  30  QAALWKDDA'                 QAALWKENA'                   GTVVAIPVE               CGTVVAILVE                250  PDTPIKAKLD	QVKKGDLIAE
40 45 50	ORF85ng and O orf85ng orf85-1 orf85ng orf85-1 orf85ng orf85-1	90 INSTTOT INSTSOT  150 ALAAAKA I:     AFAAAKA  210 PTIVQLA	W 96.1% ide  40  TVRRGDISRTV  100  NTIDMEKSKLE    ::       NTLNTEKSKLE  40  160  NVAELKALIRO  100  1  220  NLDMMLNKMO	ntity in 334  50 SATGEISPSNI  110 TYQAKLVSAQI                         TYQAKLVSAQI 50 270 SKISINTAESE                       SKISINTAESE                     230 AEGDITKVKAG	aa overlap:  60 VSVGAQASGQI             VSVGAQASGQI 10  120 1 ALGSAEKKYKE             ALGSAEKKYKE 10 180 1 DLGYTRITATME            CLGYTRITATME 10 130 240 2 CQDISFTILSEE	70  KKLYVKLGQ(                 IKILYVKLG( 20  30  QAALWKDDA'                 QAALWKENA'                GTVVAIPVE              SGTVVAILVE              PDTPIKAKLD	QVKKGDLIAE              QQVKKGDLIAE              QQVKKGDLIAE               TSKEDLESAQD                TSKEDLESAQD                COO  COO  COO  COO  COO
40 45 50	ORF85ng and O orf85ng orf85-1 orf85ng orf85-1 orf85ng orf85-1	90 INSTTOT INSTSOT  150 ALAAAKA I:IIIII AFAAAKA  210 PTIVQLA IIIIIII	W 96.1% ide  40  TVRRGDISRTV  100  NTIDMEKSKLE    ::        NTLNTEKSKLE  40  160  NVAELKALIRO  100  1  220  NLDMMLNKMQI              NLDMMLNKMQI  160  1	ntity in 334  50 SATGEISPSNI  110 TYQAKLVSAQI                     TYQAKLVSAQI 50  170 SKISINTAESE                   SKISINTAESE 10  12  230 AEGDITKVKAC 1                 AEGDITKVKAC	aa overlap:  60  VSVGAQASGQI               VSVGAQASGQI  10  120  120  1 ALGSAEKKYKR  1            ALGSAEKKYKR  1            ALGSAEKKYKR  2   10  180  1             1           1            2	70  KKLYVKLGQ(                IKILYVKLG(    20  30 QAALWKDDA'                QAALWKENA' 80  90  GTVVAIPVE              GTVVAILVE     14  250  PDTPIKAKLD              PDTPIKAKLD     20	QVKKGDLIAE              QQVKKGDLIAE 0 30  140 TSKEDLESAQD              TSKEDLESAQD 0 90  200 EGQTVNAAQST             EGQTVNAAQST 0 150  260 SVDPGLTTMSS
40 45 50	ORF85ng and O orf85ng orf85-1 orf85ng orf85-1 orf85ng orf85-1 orf85ng orf85-1	90 INSTTOT INSTSOT  150 ALAAAKA I:IIIII AFAAAKA  210 PTIVQLA IIIIIIII PTIVQLA	W 96.1% ide  40  TVRRGDISRTV  100  NTIDMEKSKLE   ::       NTLNTEKSKLE 40  160  NVAELKALIRO             NVAELKALIRO 100  220  NLDMMLNKMOI            NLDMMLNKMOI 160  280	ntity in 334  50 SATGEISPSNI  110 TYQAKLVSAQI	aa overlap:  60  VSVGAQASGQI               VSVGAQASGQI  10  120  120  1ALGSAEKKYKE                ALGSAEKKYKE  10  180  10  150  160  170  240  240  240  240  240  240  240  2	70  KKLYVKLGQQ                  IKILYVKLGQ  30  QAALWKDDA'                  QAALWKENA'  80  GTVVAIPVE               GTVVAILVE  COTPIKAKLD               COTPIKAKLD                  COTPIKAKLD                  COTPIKAKLD                  COTPIKAKLD	QVKKGDLIAE             QQVKKGDLIAE 0 30  140 TSKEDLESAQD             TSKEDLESAQD 0 90  200 EGQTVNAAQST             EGQTVNAAQST 0 150  260 SVDPGLTTMSS            SVDPGLTTMSS 0 210  320
40 45 50	ORF85ng and O orf85ng orf85-1 orf85ng orf85-1 orf85ng orf85-1	90 INSTTOT INSTSOT  150 ALAAAKA I:      AFAAAKA  210 PTIVOLA        PTIVOLA 270 GGYNSST	W 96.1% ide  40  TVRRGDISRTV  100  NTIDMEKSKLE    ::       NTLNTEKSKLE  40  160  NVAELKALIRO               NVAELKALIRO  100  220  NLDMMLNKMON               NLDMMLNKMON  160  280  DTASNAVYYYA	110 TYQAKLVSAQI 1111111111 TYQAKLVSAQI 1111111111 TYQAKLVSAQI 1111111111 SKISINTAESE 10 230 AEGDITKVKAC 1111111111 AEGDITKVKAC 70 18	aa overlap:  60 VSVGAQASGQI            VSVGAQASGQI 10  120 120 1ALGSAEKKYKE             ALGSAEKKYKE 0 70  180 1SO	70  KKLYVKLGO(                    IKILYVKLGO( 20  30  QAALWKDDA:                  QAALWKENA'  80  90  GTVVAIPVEI                GTVVAILVEI  250  PDTPIKAKLD                  PDTPIKAKLD   20  810  EIDGVKNVLL	QVKKGDLIAE              QQVKKGDLIAE              QQVKKGDLIAE               TSKEDLESAQD                TSKEDLESAQD                EGQTVNAAQST               EGQTVNAAQST               SVDPGLTTMSS               SVDPGLTTMSS               320  IPSLTVKNRGG
40 45 50	ORF85ng and O orf85ng orf85-1 orf85ng orf85-1 orf85ng orf85-1 orf85ng orf85-1	90 INSTTOT INSTSOT  150 ALAAAKA I:      AFAAAKA  210 PTIVOLA         PTIVOLA	W 96.1% ide  40  TVRRGDISRTV  100  NTIDMEKSKLE    ::      NTLNTEKSKLE  40  160  NVAELKALIRO              NVAELKALIRO  100  220  NIDMMLNKMQI              NLDMMLNKMQI               280  DTASNAVYYYA	110 TYQAKLVSAQI 111111111111111111111111111111111111	aa overlap:  60 VSVGAQASGQI             VSVGAQASGQI 10  120	70  KKLYVKLGO(                    IKILYVKLGO( 20  30  QAALWKDDA:                  QAALWKENA:   80  90  GTVVAIPVEI   1            GTVVAILVEI   1            CTPIKAKLD:   20  310  CIDGVKNVLL	QVKKGDLIAE             QQVKKGDLIAE 0 30  140 TSKEDLESAQD             TSKEDLESAQD 0 90  200 EGQTVNAAQST             EGQTVNAAQST             SVDPGLTTMSS             SVDPGLTTMSS 0 210  320 IPSLTVKNRGG
40 45 50	ORF85ng and O orf85ng orf85-1 orf85ng orf85-1 orf85ng orf85-1 orf85ng orf85-1	90 INSTTOT INSTSOT  150 ALAAAKA I:      AFAAAKA  210 PTIVOLA         PTIVOLA	W 96.1% ide  40  TVRRGDISRTV  100  NTIDMEKSKLE  !!::!!!!!  NTLNTEKSKLE  40  160  NVAELKALIRO  100  1  220  NLDMMLNKMQI  !!!!!!!!!  NLDMMLNKMQI  160  280  DTASNAVYYYA  !!!!!!!!!!	110 TYQAKLVSAQI 1110 TYQAKLVSAQI 11111111111 TYQAKLVSAQI 50 170 SKISINTAESE 11111111111 SKISINTAESE 110 230 AEGDITKVKAQ 1111111111 AEGDITKVKAQ 290 ARSFVPNPDGKI ARSFVPNPDGKI	aa overlap:  60 VSVGAQASGQI             VSVGAQASGQI 10  120	70  KKLYVKLGQQ                  IKILYVKLGQ  20  30  QAALWKDDA:                 QAALWKENA:   80  90  GTVVAIPVE               GTVVAILVE   250  PDTPIKAKLD:              PDTPIKAKLD:                 SIDGVKNVLL                 SIDGVKNVLL	QVKKGDLIAE              QQVKKGDLIAE              QQVKKGDLIAE               TSKEDLESAQD                TSKEDLESAQD                EGQTVNAAQST                EGQTVNAAQST                SVDPGLTTMSS                SVDPGLTTMSS                320  IPSLTVKNRGG                 IPSLTVKNRGG
40 45 50	ORF85ng and O orf85ng orf85-1 orf85ng orf85-1 orf85ng orf85-1 orf85ng orf85-1	90 INSTTOT INSTSOT  150 ALAAAKA I:IIIII AFAAAKA 210 PTIVOLA IIIIIIII PTIVOLA 270 GGYNSST	W 96.1% ide  40  TVRRGDISRTV  100  NTIDMEKSKLE  !!::!!!!!  NTLNTEKSKLE  40  160  NVAELKALIRO  100  1  220  NLDMMLNKMQI  !!!!!!!!!!  NLDMMLNKMQI  160  280  DTASNAVYYYA  !!!!!!!!!!	110 TYQAKLVSAQI 1110 TYQAKLVSAQI 11111111111 TYQAKLVSAQI 50 170 SKISINTAESE 11111111111 SKISINTAESE 110 230 AEGDITKVKAQ 1111111111 AEGDITKVKAQ 290 ARSFVPNPDGKI ARSFVPNPDGKI	aa overlap:  60 VSVGAQASGQI             VSVGAQASGQI 10  120	70  KKLYVKLGQQ                  IKILYVKLGQ  20  30  QAALWKDDA:                 QAALWKENA:   80  90  GTVVAIPVE               GTVVAILVE   250  PDTPIKAKLD:              PDTPIKAKLD:                 SIDGVKNVLL                 SIDGVKNVLL	QVKKGDLIAE             QQVKKGDLIAE 0 30  140 TSKEDLESAQD             TSKEDLESAQD 0 90  200 EGQTVNAAQST             EGQTVNAAQST             SVDPGLTTMSS             SVDPGLTTMSS 0 210  320 IPSLTVKNRGG
40 45 50 55	ORF85ng and O orf85ng orf85-1 orf85ng orf85-1 orf85ng orf85-1 orf85ng orf85-1	90 INSTTOT INSTSOT  150 ALAAAKA I:      AFAAAKA  210 PTIVOLA         PTIVOLA	W 96.1% ide  40  TVRRGDISRTV  100  NTIDMEKSKLE  !!::!!!!!  NTLNTEKSKLE  40  160  NVAELKALIRO  100  1  220  NLDMMLNKMQI  !!!!!!!!!!  NLDMMLNKMQI  160  280  DTASNAVYYYA  !!!!!!!!!!	110 TYQAKLVSAQI 1110 TYQAKLVSAQI 11111111111 TYQAKLVSAQI 50 170 SKISINTAESE 11111111111 SKISINTAESE 110 230 AEGDITKVKAQ 1111111111 AEGDITKVKAQ 290 ARSFVPNPDGKI ARSFVPNPDGKI	aa overlap:  60  VSVGAQASGQI             VSVGAQASGQI 10  120	70  KKLYVKLGQ(                 IKILYVKLG( 20  30  QAALWKDDA:                 QAALWKENA:   80  90  GTVVAIPVE:                 POTPIKAKLD:                 POTPIKAKLD:                 POTPIKAKLD:                 POTPIKAKLD:                 EIDGVKNVLL:                 EIDGVKNVLL:	QVKKGDLIAE              QQVKKGDLIAE              QQVKKGDLIAE               TSKEDLESAQD                TSKEDLESAQD                EGQTVNAAQST                EGQTVNAAQST                SVDPGLTTMSS                SVDPGLTTMSS                320  IPSLTVKNRGG                 IPSLTVKNRGG

orf85ng

-431-

KAFVRVLGADGKAVEREIRTGMKDSMNTEVKSGLKEGDKVVISEITAAEQQESGERALGG

```
orf85-1
                       KAFVRVLGADGKAAEREIRTGMRDSMNTEVKSGLKEGDKVVISEITAAEQQESGERALGG
                                                  300
                              280
                                        290
                                                            310
5
                       390
          orf85ng
                        PPRRX
                        HHH
          orf85-1
                        PPRRX
     In addition, ORF85ng shows significant homology to an E.coli membrane fusion protein:
10
          gi|1787104 (AE000189) o380; 27% identical (27 gaps) to 332 residues from
          membrane fusion protein precursor, MTRC_NEIGO SW: P43505 (412 aa) [Escherichia
          coli] Length = 380
           Score = 193 \text{ bits } (485), \text{ Expect = } 2e-48
15
            Identities = 120/345 (34%), Positives = 182/345 (51%), Gaps = 13/345 (3%)
                     PQAAYITETVRRGDISRTVSATGEISPSNLVSVGAQASGQIKKLYVKLGQQVKKGDLIAE 88
                         Y T VR GD+ ++V ATG++
                                                   V VGAQ SGQ+K L V +G +VKK L+
           Sbict: 41
                     PVPTYQTLIVRPGDLQQSVLATGKLDALRKVDVGAQVSGQLKTLSVAIGDKVKKDQLLGV 100
20
                      INSTTQTNTIDMEKSKLETYQAKLVSAQIALGSAEKKYKRQAALWKDDATSKEXXXXXXX 148
                            N I ++ L +A+ A+ L A Y RQ L + A S++
           Sbjct: 101 IDPEQAENQIKEVEATLMELRAQRQQAEAELKLARVTYSRQQRLAQTKAVSQQDLDTAAT 160
25
           Query: 149 XXXXXXXXXXXXXXXXIRQSKISINTAESDLGYTRITATMDGTVVAIPVEEGQTVNAAQST 208
                                    I++++ S++TA+++L YTRI A M G V I
                                                                     +GQTV AAQ
           Sbjct: 161 EMAVKQAQIGTIDAQIKRNQASLDTAKTNLDYTRIVAPMAGEVTQITTLQGQTVIAAQQA 220
           Query: 209 PTIVQLANLDMMLNKMQIAEGDITKVKAGQDISFTILSEPDTPIKAKLDSVDPGLTTMSS 268
30
                      P I+ LA++ ML K Q++E D+ +K GQ
                                                     FT+L +P T + ++ V P
           Sbjct: 221 PNILTLADMSAMLVKAQVSEADVIHLKPGQKAWFTVLGDPLTRYEGQIKDVLP----- 273
           Query: 269 GGYNSSTDTASNAVYYYARSFVPNPDGKLATGMTTQNTVEIDGVKNVLLIPSLTVKNRGG 328
                           + + ++A++YYAR VPNP+G L
                                                    MT Q +++ VKNVL IP
35
           Sbjct: 274 ----TPEKVNDAIFYYARFEVPNPNGLLRLDMTAQVHIQLTDVKNVLTIPLSALGDPVG 328
           Query: 329 KAFVRV-LGADGKAVEREIRTGMKDSMNTEVKSGLKEGDKVVISE 372
                          +V L +G+ ERE+ G ++ + E+ GL+ GD+VVI E
           Sbjct: 329 DNRYKVKLLRNGETREREVTIGARNDTDVEIVKGLEAGDEVVIGE 373
40
      Based on this analysis, it was predicted that the proteins from N.meningitidis and N.gonorrhoeae,
      and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.
      ORF85-1 (40.4kDa) was cloned in the pGex vectors and expressed in E.coli, as described above.
      The products of protein expression and purification were analyzed by SDS-PAGE. Figure 19A
```

#### Example 92

45

The following partial DNA sequence was identified in N.meningitidis <SEO ID 773>:

shows the results of affinity purification of the GST-fusion protein. Purified GST-fusion protein

was used to immunise mice, whose sera were used for Western blot (Figure 19B), FACS analysis

(Figure 19C), and ELISA (positive result). These experiments confirm that ORF85-1 is a

```
50 1 .ATTCCCGCCA CGATGACATT TGAACGCAGC GGCAATGCTT ACAAAATCGT
51 TTCGACGATT AAAGTGCCGC TATACAATAT CCGTTTCGAG TCCGGCGGTA
101 CGGTTGTCGG CAATACCCTG CACCCTACCT ACTATAGAGA CATACGCAGG
151 GGCAAACTGT ATGCGGAAGC CAAATTCGCC GACGGCAGCG TAACTTACGG
201 CAAAGCGGGC GAGAGCAAAA CCGAGCAAAG CCCCAAGGCT ATGGATTTGT
```

surface-exposed protein, and that it is a useful immunogen.

5	TCACGCTTGC CTGGCAGTTG GCGGCAAATT ACGCGAAACT CCCCCCGGGG  TCACAAATCA CCAACGGCAA AAAACTTTAT TCCGTCGGCG GTTTGAATAA  TCACGGTACA GGAAAATACA GCATAGGCGG CGTGGAAACC GAAGTCGTCA  AATATCGGGT GCGGCGGGC GACGATGCGG TAATGTATTT CTTCGCACCG  TCCCTGAACA ATATTCCGGC ACAAATCGGC TATACCGACG ACGCCAAAC  TCATACGCTG AAACTCAAAT CGGTGCAGAT CAACGGCCAG GCAGCCAAAC  CGTAA
	This corresponds to the amino acid sequence <seq 774;="" id="" orf120="">:</seq>
10	1IPATMTFERS GNAYKIVSTI KVPLYNIRFE SGGTVVGNTL HPTYYRDIRR 51 GKLYAEAKFA DGSVTYGKAG ESKTEQSPKA MDLFTLAWQL AANDAKLPPG 101 LKITNGKKLY SVGGLNKAGT GKYSIGGVET EVVKYRVRRG DDAVMYFFAP 151 SLNNIPAQIG YTDDGKTYTL KLKSVQINGQ AAKP*
	Further work revealed the complete nucleotide sequence <seq 775="" id="">:</seq>
15	1 ATGATGAAGA CTTTTAAAAA TATATTTTCC GCCGCCATTT TGTCCGCCGC 51 CCTGCCGTGC GCGTATGCGG CAGGGCTGCC CCAATCCGCC GTGCTGCACT 101 ATTCCGGCAG CTACGGCATT CCCGCCACGA TGACATTTGA ACGCAGCGGC 151 AATGCTTACA AAATCGTTTC GACGATTAAA GTGCCGCTAT ACAATATCCG 201 TTTCGAGTCC GGCGGTACGG TTGTCGGCAA TACCCTGCAC CCTACCTACT
20	251 ATAGAGACAT ACGCAGGGGC AAACTGTATG CGGAAGCCAA ATTCGCCGAC 301 GGCAGCGTAA CTTACGGCAA AGCGGGCGAG AGCAAAACCG AGCAAAGCCC 351 CAAGGCTATG GATTTGTTCA CGCTTGCCTG GCAGTTGGCG GCAAATGACG 401 CGAAACTCCC CCCGGGGCTG AAAATCACCA ACGGCAAAAA ACTTTATTCC 451 GTCGGCGGTT TGAATAAGGC GGGTACAGGA AAATACAGCA TAGGCGGCGT
25	501 GGAAACCGAA GTCGTCAAAT ATCGGGTGCG GCGCGGCGAC GATGCGGTAA 551 TGTATTTCTT CGCACCGTCC CTGAACAATA TTCCGGCACA AATCGGCTAT 601 ACCGACGACG GCAAAACCTA TACGCTGAAA CTCAAATCGG TGCAGATCAA 651 CGGCCAGGCA GCCAAACCGT AA
	This corresponds to the amino acid sequence <seq 776;="" id="" orf120-1="">:</seq>
30	1 MMKTFKNIFS AAILSAALPC AYAAGLPQSA VLHYSGSYGI PATMTFERSG 51 NAYKIVSTIK VPLYNIRFES GGTVVGNTLH PTYYRDIRRG KLYAEAKFAD 101 GSVTYGKAGE SKTEQSPKAM DLFTLAWQLA ANDAKLPPGL KITNGKKLYS 151 VGGLNKAGTG KYSIGGVETE VVKYRVRRGD DAVMYFFAPS LNNIPAQIGY 201 TDDGKTYTLK LKSVQINGQA AKP*
	Computer analysis of this amino acid sequence gave the following results:
35	Homology with a predicted ORF from N.meningitidis (strain A)
	ORF120 shows 92.4% identity over a 184aa overlap with an ORF (ORF120a) from strain A of N.
	meningitidis:
	10 20 30
40	orf120.pep IPATMTFERSGNAYKIVSTIKVPLYNIRFE
45	40 50 60 70 80 90 orf120.pep SGGTVVGNTLHPTYYRDIRRGKLYAEAKFADGSVTYGKAGESKTEQSPKAMDLFTLAWQL
	orf120a SGGTVVGNTLHPTYYRDIRRGKLYAEAKFADGSVTYGKAXXXXXXQSPKAMDLFTLAWQL 70 80 90 100 110 120
50	100 110 120 130 140 150 orf120.pep AANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETEVVKYRVRRGDDAVMYFFAP
55	orf120a AANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETEVVKYRVRRGDDAVMYFFAP 130 140 150 160 170 180

		160 170 180
		SLNNIPAQIGYTDDGKTYTLKLKSVQINGQAAKPX
5		SLNN I PAQIGYTDDGKTYTLKLKSVQINGQAAKPX
•		ORF120a nucleotide sequence <seq 777="" id=""> is:</seq>
	_	ATGAAGA CTTTTAAAAA TATATTTTCC GCCGCCATTT TGTCCGCCGC
	51 CCT	GCCGTGC GCGTATGCGG CAGGGCTGCC CNAATCCGCC GTGCTGCACT CCGGCAG CTACGGCATT CCCGCCACNA NNANNTNNGN ACNNNGNGNC
10	151 AAT	GCTTNCA AAATCGTTTC GACGATTAAA GTGCCGCTAT ACAATATCCG
		CGAGTCC GGCGGTACGG TTGTCGGCAA TACCCTGCAC CCTACCTACT GAGACAT ACGCAGGGGC AAACTGTATG CGGAAGCCAA ATTCGCCGAC
	301 GGC	AGCGTAA CCTACGGCAA AGCGGNNNNN ANCNNNNNNG NGCAAAGCCC
1.5	351 CAA	GGCTATG GATTTGTTCA CGCTTGCNTG GCAGTTGGCG GCAAATGACG AACTCCC CCCGGGGCTG AAAATCACCA ACGGCAAAAA ACTTTATTCC
15	401 CGA 451 GTC	GGCGGTT TGAATAAGGC GGGTACAGGA AAATACAGCA TAGGCGGCGT
	501 GGA	AACCGAA GTCGTCAAAT ATCGGGTGCG GCGCGGCGAC GATGCGGTAA
	551 TGT 601 ACC	ATTTCTT CGCACCGTCC CTGAACAATA TTCCGGCACA AATCGGCTAT GACGACG GCAAAACCTA TACGCTGAAA CTCAAATCGG TGCAGATCAA
20	651 CGG	CCAGGCA GCCAAACCGT AA
	This encodes a prote	in having amino acid sequence <seq 778="" id="">:</seq>
	1 MMK	TFKNIFS AAILSAALPC AYAAGLPXSA VLHYSGSYGI PATXXXXXXX
	51 NAX	KIVSTIK VPLYNIRFES GGTVVGNTLH PTYYRDIRRG KLYAEAKFAD
25		TYGKAXX XXXXQSPKAM DLFTLAWQLA ANDAKLPPGL KITNGKKLYS LNKAGTG KYSIGGVETE VVKYRVRRGD DAVMYFFAPS LNNIPAQIGY
23		OGKTYTLK LKSVQINGQA AKP*
	ORF120a and ORF1	20-1 show 93.3% identity in 223 aa overlap:
		10 20 30 40 50 60
20	orf120a.pep	MMKTFKNIFSAAILSAALPCAYAAGLPXSAVLHYSGSYGIPATXXXXXXXNAXKIVSTIK
30	orf120-1	
	011110	10 20 30 40 50 60
		70 80 90 100 110 120
35	orf120a.pep	VPLYNIRFESGGTVVGNTLHPTYYRDIRRGKLYAEAKFADGSVTYGKAXXXXXXQSPKAM
	orf120-1	VPLYNIRFESGGTVVGNTLHPTYYRDIRRGKLYAEAKFADGSVTYGKAGESKTEQSPKAM
		70 80 90 100 110 120
40		130 140 150 160 170 180
	orf120a.pep	DLFTLAWQLAANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETEVVKYRVRRGD
	orf120-1	DLFTLAWQLAANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETEVVKYRVRRGD
45		130 140 150 160 170 180
72		190 200 210 220
	orf120a.pep	DAVMYFFAPSLNNIPAQIGYTDDGKTYTLKLKSVQINGQAAKPX 
	orf120-1	DAVMYFFAPSLNNIPAQIGYTDDGKTYTLKLKSVQINGQAAKPX
50		190 200 210 220
	Homology with a p	redicted ORF from N.gonorrhoeae
	ORF120 shows 97.	.8% identity over 184 aa overlap with a predicted ORF (ORF120ng) from
	N.gonorrhoeae:	
55	orf120.pep	IPATMTFERSGNAYKIVSTIKVPLYNIRFE 30
	orf120ng	SAAILSAALPCAYAARLPQSAVLHYSGSYGIPATMTFERSGNAYKIVSTIKVPLYNIRFE 69
<b>~</b> ^	orf120.pep	SGGTVVGNTLHPTYYRDIRRGKLYAEAKFADGSVTYGKAGESKTEQSPKAMDLFTLAWQL 90
60	orf120ng	
	022220119	

orf120.pep

-434-

AANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETEVVKYRVRRGDDAVMYFFAP 150

	OIIIZO. Pep	
	orf120ng	AANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETEVVKYRVRRGDDTVTYFFAP 189
5	, , , , , , , , , , , , , , , , , , ,	
	orf120.pep	SLNNIPAQIGYTDDGKTYTLKLKSVQINGQAAKP 184
	orf120ng	
	The complete length	n ORF120ng nucleotide sequence <seq 779="" id=""> is:</seq>
10		
10		GATGAAGA CTTTTAAAAA TATATTTTCC GCCGCCATTT TGTCCGCCGC
		FGCCGTGC GCGTATGCGG CAAGGCTACC CCAATCCGCC GTGCTGCACT FCCGGCAG CTACGGCATT CCCGCCACGA TGACATTTGA ACGCAGCGGC
	151 AA	GCTTACA AAATCGTTTC GACGATTAAA GTGCCGCTAT ACAATATCCG
	201 TT	CGAATCC GGCGGTACGG TTGTCGGCAA TACCCTGCAC CCTGCCTACT
15	251 AT	AAAGACAT ACGCAGGGGC AAACTGTATG CGGAAGCCAA ATTCGCCGAC
	301 GG	CAGCGTAA CCTACGGCAA AGCGGGCGAG AGCAAAACCG AGCAAAGCCC
	351 CA	AGGCTATG GATTTGTTCA CGCTTGCCTG GCAGTTGGCG GCAAATGACG
	401 CG/	AAACTCCC CCCGGGTCTG AAAATCACCA ACGGCAAAAA ACTTTATTCC
	451 GT	CGGCGGCC TGAATAAGGC GGGTACGGGA AAATACAGCA TaggCGGCGT
20	501 GG/	AAACCGAA GTCGTCAAAT ATCGGGTGCG GCGCGGCGAC GATACGGTAA
	551 CG	PATTTCTT CGCACCGTCC CTGAACAATA TTCCGGCACA AATCGGCTAT
		CGACGACG GCAAAACCTA TACGCTGAAG CTCAAATCGG TGCAGATCAA
	651 CG	GACAGGCC GCCAAACCGT AA
	This encodes a prote	ein having amino acid sequence <seq 780="" id="">:</seq>
	This cheddes a prod	cin having animo acid sequence SEQ ID 780%.
25	1 MM	KTFKNIFS AAILSAALPC AYAARLPOSA VLHYSGSYGI PATMTFERSG
23		YKIVSTIK VPLYNIRFES GGTVVGNTLH PAYYKDIRRG KLYAEAKFAD
		VTYGKAGE SKTEQSPKAM DLFTLAWQLA ANDAKLPPGL KITNGKKLYS
	151 VG	GLNKAGTG KYSIGGVETE VVKYRVRRGD DTVTYFFAPS LNNIPAQIGY
		DGKTYTLK LKSVQINGQA AKP*
30	In comparison with	ORF120-1, ORF120ng shows 97.8% identity in 223 aa overlap:
	5100 1	10 20 30 40 50 60
	orf120-1.pep	
	orf120ng	
35	OFFIZONG	
33	•	10 20 30 40 50 60
		70 80 90 100 110 120
	orf120-1.pep	
40	orf120ng	VPLYNIRFESGGTVVGNTLHPAYYKDIRRGKLYAEAKFADGSVTYGKAGESKTEQSPKAM
		70 80 90 100 110 120
	6300 1	130 140 150 160 170 180
45	orf120-1.pep	
43	orf120ng	
	Offizong	130 140 150 160 170 180
		130 140 150 100 170 160
		190 200 210 220
50	orf120-1.pep	DAVMYFFAPSLNNIPAQIGYTDDGKTYTLKLKSVQINGQAAKPX
	orf120ng	DTVTYFFAPSLNNIPAQIGYTDDGKTYTLKLKSVQINGQAAKPX
		190 200 210 220
	This ample de de de	Alima the massenge of a mutation leader to the state of t
	i nis analysis, incli	iding the presence of a putative leader sequence in the gonococcal protein

This analysis, including the presence of a putative leader sequence in the gonococcal protein suggests that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

## Example 93

55

The following partial DNA sequence was identified in N.meningitidis <SEQ ID 781>:

orf121.pep

```
1 ATGTATCGGA GGAAAGGGCG GGGCATCAAG CCGTGGATGG GTGCCGGTGC
                      .GCGTTTGCC GCCTTGGTCT GGCTGGTTTT CGCGCTCGGC GATACTTTGA
                 51
                     CTCCGTTTGC GGTTGCGGCG GTGCTGGCGT ATGTATTGGA CCCTTTGGTC
                101
                151
                     GAATGGTTGC AGAAAAAGGG TTTGAACCGT GCATCCGCTT CGATGTCTGT
5
                     GATGGTGTTT TCCTTGATTT TGTTGTTGGC ATTATTGTTG ATTATCGTCC
                201
                251
                     CTATGCTGGT CGGGCAGTTC AACAATTTGG CATCGCGCCT GCCCCAATTA
                     ATCGGTTTTA TGCAGAACAC GCTGCTGCCG TGGTTGAAAA ATACAATCGG
                301
                351
                     CGGATATGTG GAAATCGATC AGGCATCTAT TATTGCGTGG CTTCAGGCGC
                     ATACGGGAGA GTTGAGCAAC GCGCTTAAGG CGTGGTTTCC CGTTTTGATG
                401
10
                451 AGGCAGGGCG GCAATATT..
     This corresponds to the amino acid sequence <SEQ ID 782; ORF121>:
                  1 MYRRKGRGIK PWMGAGXAFA ALVWLVFALG DTLTPFAVAA VLAYVLDPLV
                 51 EWLQKKGLNR ASASMSVMVF SLILLLALLL IIVPMLVGQF NNLASRLPQL
                101 IGFMONTLLP WLKNTIGGYV EIDQASIIAW LQAHTGELSN ALKAWFPVLM
15
                151 ROGGNI..
     Further work revealed the complete nucleotide sequence <SEQ ID 783>:
                   1 ATGTATCGGA GGAAAGGCG GGGCATCAAG CCGTGGATGG GTGCCGGTGC
                      GGCGTTTGCC GCCTTGGTCT GGCTGGTTTT CGCGCTCGGC GATACTTTGA
                  51
                101
                      CTCCGTTTGC GGTTGCGGCG GTGCTGGCGT ATGTATTGGA CCCTTTGGTC
20
                      GAATGGTTGC AGAAAAAGGG TTTGAACCGT GCATCCGCTT CGATGTCTGT
                151
                      GATGGTGTTT TCCTTGATTT TGTTGTTGGC ATTATTGTTG ATTATCGTCC CTATGCTGGT CGGGCAGTTC AACAATTTGG CATCGCGCCT GCCCCAATTA
                201
                251
                      ATCGGTTTTA TGCAGAACAC GCTGCTGCCG TGGTTGAAAA ATACAATCGG
                301
                      CGGATATGTG GAAATCGATC AGGCATCTAT TATTGCGTGG CTTCAGGCGC
                 351
25
                      ATACGGGAGA GTTGAGCAAC GCGCTTAAGG CGTGGTTTCC CGTTTTGATG
                 401
                      AGGCAGGGCG GCAATATTGT CAGCAGTATC GGCAACCTGC TGCTGCTTCC
                 451
                      CTTGCTGCTT TACTATTTCC TGCTGGATTG GCAGCGGTGG TCGTGCGGCA
                 501
                      TTGCCAAACT GGTTCCGAGG CGTTTTGCCG GTGCTTATAC GCGCATTACA
                 551
                      GGCAATTTGA ACGAGGTATT GGGCGAATTT TTGCGCGGGC AGCTTCTGGT
                 601
                      AATGCTGATT ATGGGCTTGG TTTACGGTTT GGGATTGGTG CTGGTCGGGC
30
                 651
                      TGGATTCGGG GTTTGCCATC GGTATGCTTG CCGGTATTTT GGTGTTTGTC
                 701
                      CCTTATCTCG GGGCGTTTAC GGGATTGCTG CTTGCCACCG TCGCCGCCTT
                 751
                      GCTCCAGTTC GGTTCGTGGA ACGGCATCCT ATCGGTTTGG GCGGTTTTTG
                 801
                      CCGTAGGACA GTTTCTCGAA AGTTTTTTCA TTACGCCGAA AATCGTGGGA
                 851
35
                 901
                      GACCGTATCG GGCTGTCGCC GTTTTGGGTT ATCTTTTCGC TGATGGCGTT
                      CGGGCAGCTG ATGGGCTTTG TCGGAATGTT GGCGGGATTG CCTTTGGCCG
                 951
                      CCGTAACCTT GGTCTTGCTT CGCGAGGGCG TGCAGAAATA TTTTGCCGGC
                1001
                1051 AGTTTTTACC GGGGCAGGTA G
      This corresponds to the amino acid sequence <SEQ ID 784; ORF121-1>:
                      MYRRKGRGIK PWMGAGAAFA ALVWLVFALG DTLTPFAVAA VLAYVLDPLV EWLQKKGINR ASASMSVMVF SLILLLALLL IIVPMLVGQF NNLASRLPQL
40
                  51
                 101
                      IGFMONTLLP WLKNTIGGYV EIDQASIIAW LQAHTGELSN ALKAWFPVLM
                      ROGGNIVSSI GNLLLLPLLL YYFLLDWQRW SCGIAKLVPR RFAGAYTRIT
                 151
                      GNLNEVLGEF LRGQLLVMLI MGLVYGLGLV LVGLDSGFAI GMLAGILVFV
PYLGAFTGLL LATVAALLQF GSWNGILSVW AVFAVGQFLE SFFITPKIVG
                 201
 45
                 251
                      DRIGLSPFWV IFSLMAFGQL MGFVGMLAGL PLAAVTLVLL REGVQKYFAG
                 301
                      SFYRGR*
       Computer analysis of this amino acid sequence gave the following results:
       Homology with a predicted ORF from N.meningitidis (strain A)
       ORF121 shows 98.7% identity over a 156aa overlap with an ORF (ORF121a) from strain A of N.
 50
       meningitidis:
                                             20
                                                       30
                                                                  40
                          MYRRKGRGIKPWMGAGXAFAALVWLVFALGDTLTPFAVAAVLAYVLDPLVEWLQKKGLNR
            orf121.pep
                          55
                          MYRRKGRGIKPWMDAGAAFAALVWLVFALGDTLTPFAVAAVLAYVLDPLVEWLQKKGLNR
            orf121a
                                                                            50
                                  10
                                             20
                                                       30
                                                                  40
                                             80
                                                       90
                                                                 100
                                                                           110
                                                                                      120
```

ASASMSVMVFSLILLLALLLIIVPMLVGOFNNLASRLPQLIGFMQNTLLPWLKNTIGGYV

	orf121a	ASASMSVM	/FSLILLLALL		NNLASRLPQL 100		WLKNTIGGY	V
5	orf121.pep	EIDQASIIA	_	nalkawfpvli	<b>IRQGGNI</b>		;	
	orf121a	EIDQASII	AWLQAHTGELS		ARQGGNIVSSI			
10			30 14					
	orf121a		PRRFAGAYTRI 90 20		FLRGQLLVMLI 220			
	The complete lengt	h ORF121	a nucleotide	sequence <	SEQ ID 785	i> is:		
15					CCGTGGATGG CGCGCTCGGC			
15					ATGTATTGGA			
	151 GA	ATGGTTGC	AGAAAAAGGG	TTTGAACCGT	GCATCCGCTT	CGATGTCTC	<b>ST</b>	
					ATTATTGTTG CATCGCGCCT			
20					TGGTTGAAAA			
~~	•				TATTGCGTGG			
					CGTGGTTTCC			
					GGCAACCTGC			
25					GCAGCGGTGG GTGCTTATAC			
, 23					TTGCGCGGGC			
					GGGGTTGGTG			
					CCGGTATTT			
20					CTGGCAACCG			
30					GGCTGTTTGG TTACGCCGAA			
					ATCTTTTCGC			
					GGCCGGATTG			
					TGCAGAAATA	TTTTGCCG	GC	
35			GGGGCAGGTA		∙CEΛ ID 704	<u>~.</u>		
	This encodes a pro-	iciii naviii	g ammo acio	sequence >	SEQ ID 180	<b>)</b> .		
	1 M	RRKGRGIK	PWMDAGAAFA	ALVWLVFALG	DTLTPFAVAA	VLAYVLDP	LV	
	51 EV	VLQKKGLNR	ASASMSVMVF	SLILLLALLL	IIVPMLVGQE	NNLASRLP	QL	
40					LQAHTGELSN			
40					SCGIAKLVPF			
					LVGLDSGFAI AVFAVGQFLE			
	301 D	RIGLSPFWV			PLAAVTLVLI			
	501 5.	FYRGR*						
45	ORF121a and ORI	F121-1 sho	w99.2% ide	ntity in 356	aa overlap:			
	· 		10	20	30	40	50	60
	orf121a.pep				ALGOTLTPFAV		-	
	orf121-1							
50	011121 1	1111444	10	20	30	40	50	60
			70	80	90 1	100	110	120
	orf121a.pep	ASASMS			GOFNNLASRLI		110 T.T.PWI.KNTT	
	01111101505							
55	orf121-1	ASASMS			GOFNNLASRL			
	•		70	80	90 :	100	110	120
			130	140		160	170	180
<i>6</i> 0	orf121a.pep				PVLMRQGGNIV			
60	orf121-1							
	. 011121-1	nināu	130	140		160	170	180
65	A-6101	COCTAT	190	200		220 Mi thetavet	230	240
ΟJ	orf121a.pep	SUGIAL	VTA EKKEAPAA	IKTIGNUNEVI	LGEFLRGQLLV	MTTMOTA IRT	GLAFAGFD2	<b>GFAI</b>

PCT/IB98/01665

WO 99/24578 -437-

	orf121-1	SCGIAKLVPRRFAGA			  LLVMLIMGLV   220	 YGLGLVLVG: 230	 LDSGFAI 240
5		250	260	270	280	290	300
5	orf121a.pep	GMVAGILVFVPYLG	AFTGLLLAT	VAALLQFGSWN	<b>IGILAVWAVFA</b>	VGQFLESFF	ITPKIVG
	Offizia. bob	11:1111111111		нийии		111111111	1111111
	orf121-1	GMLAGILVFVPYLG	AFTGLLLAT	VAALLOFGSWN	IGILSVWAVFA	VGQFLESFF	ITPKIVG
	OIIIZI-I	250	260	270	280	290	300
10	•	200					
10		310	320	330	340	350	
	orf121a.pep	DRIGLSPFWVIFSL	MAFGQLMGF	VGMLAGLPLA/	AVTLVLLREGV	<b>OKYFAGSFY</b>	RGRX
	OLLINITATION		11111111	[[[[[]]]]]		111111111	1111
	orf121-1	DRIGLSPFWVIFSL	MAFGOLMGF	VGMLAGLPLA	AVTLVLLREGV	<b>OKYFAGSFY</b>	RGRX
15	011121 1	310	320	. 330	340	350	

# Homology with a predicted ORF from N.gonorrhoeae

ORF121 shows 97.4% identity over a 156 aa overlap with a predicted ORF (ORF121ng) from

### N.gonorrhoeae:

20	orf121.pep	MYRRKGRGIKPWMGAGXAFAALVWLVFALGDTLTPFAVAAVLAYVLDPLVEWLQKKGLNR 	60 60
	orf121ng	MIKKNGKGINPWMGAGAAFAALDVWLVIALDDILIFERVAAVUALVBDELVEWBONGCOLOU	•
25	orf121.pep	ASASMSVMVFSLILLLALLLIIVPMLVGQFNNLASRLPQLIGFMQNTLLPWLKNTIGGYV	120
23	orf121ng	ASASMSVMVFSLILLLALLLIIVPMLVGQFNNLASRLPQLIGFMQNTLLPWLKNTIGGYV	120
			156
	orf121.pep	EIDQASIIAWLQAHTGELSNALKAWFPVLMRQGGNI	120
30	orf121ng		180

An ORF121ng nucleotide sequence <SEQ ID 787> was predicted to encode a protein having amino acid sequence <SEQ ID 788>:

```
MYRRKGRGIK PWMGAGAAFA ALVWLVYALG DTLTPFAVAA VLAYVLDPLV
                   1 MYRRKGRGIK PWMGAGAAFA ALVWLVYALG DTLTPFAVAA VLAYVLDPLV
51 EWLQKKGLNR ASASMSVMVF SLILLLALLL IIVPMLVGQF NNLASRLPQL
                   101 IGFMONTLLP WLKNTIGGYV EIDQASIIAW FQAHTGELSN ALKAWFPVLM
35
                   151 KQGGNIVSTI GNLLLPPLLL YYFLLDWHRW SCGIPKLVPR RFAGAYTRIT
                        GNLNKVWGKF LRGQLLGETE RGAVVCRVGR ECWEGGGARS RPSDDGWPRW
                        GGG*
                   251
```

# Further work revealed the following gonoccocal DNA sequence <SEQ ID 789>:

40	1	ATGTATCGGA	GAAAAGGACG	GGGCATCAAG	CCGTGGATGG	GTGCCGGCGC
	51	GGCGTTTGCC	GCCTTGGTCT	GGCTGGTTTA	CGCGCTCGGC	GATACTTTGA
	101	CTCCGTTTGC	GGTTGCGGCG	GTGCTGGCGT	ATGTGTTGGA	CCCTTTGGTC
	151	GAATGGTTGC	AGAAAAAGGG	TTTGAACCGT	GCATCCGCTT	CGATGTCTGT
	201	GATGGTGTTT	TCCTTGATTT	TGTTGTTGGC	ATTATTGTTG	ATTATTGTCC
45	251	CTATGCTGGT	CGGGCAGTTC	AATAATTTGG	CATCTCGCCT	GCCCCAATTA
	301	ATCGGTTTTA	TGCAGAACAC	GCTGCTGCCG	TGGTTGAAAA	
	351	CGGATATGTG	GAAATCGATC	AGGCATCTAT	TATTGCGTGG	TTTCAGGCGC
	401	ATACGGGCGA	GTTGAGCAAC	GCGCTTAAGG	CGTGGTTTCC	CGTTTTGATG
	451	AAACAGGGCG	GCAATATTGT	CAGCAGTATC	GGCAACCTGC	TGCTGCCGCC
50	501	CTTGCTGCTT	TACTATTTCC	TGCTGGATTG	GCAGCGGTGG	
	551	TCGCCAAACT	GGTTCCGAGG	CGTTTTGCCG	GTGCTTATAC	
	601	GGTAATTTGA	ACGAGGTATT	GGGCGAATTT	TTGCGCGGTC	
	651	GATGCTGATT	ATGGGCTTGG	TTTACGGTTT	GGGATTGATG	CTAGTCGGAC
	701	TGGATTCGGG	ATTTGCCATC	GGTATGGTTG		GGTGTTTGTC
55	751	CCCTATTTGG	GTGCGTTTAC	GGGATTGCTG	CTTGCCACTG	TTGCAGCCTT
	801	GCTCCAGTTC	GGTTCGTGGA	ACGGAATCTT	GGCTGTTTGG	GCGGTTTTTG
•	851	CCGTCGGTCA	GTTTCTCGAA	AGTTTTTTCA		
	901	GACCGTATCG	GCCTGTCGCC	GTTTTGGGTT	ATCTTTTCGC	
	951	CGGAGAGCTG	ATGGGCTTTG	TCGGAATGTT		
60	1001	CCGTAACCTT	GGTCTTGCTT	CGCGAGGGCG	CGCAGAAATA	TTTTGCCGGC
	1051	AGTTTTTACC	GGGGCAGGTA	G		

This corresponds to the amino acid sequence <SEQ ID 790; ORF121ng-1>:

```
MYRRKGRGIK PWMGAGAAFA ALVWLVYALG DTLTPFAVAA VLAYVLDPLV
                  EWLQKKGLNR ASASMSVMVF SLILLLALLL IIVPMLVGQF NNLASRLPQL
               51
              101
                  IGFMONTLLP WLKNTIGGYV EIDQASIIAW FQAHTGELSN ALKAWFPVLM
                  KQGGNIVSSI_GNLLLPPLLL YYFLLDWQRW SCGIAKLVPR RFAGAYTRIT
5
              151
              201
                  GNLNEVLGEF LRGQLLVMLI MGLVYGLGLM LVGLDSGFAI GMVAGILVFV
                  PYLGAFTGLL LATVAALLQF GSWNGILAVW AVFAVGQFLE SFFITPKIVG
              251
              301
                  DRIGLSPFWV IFSLMAFGEL MGFVGMLAGL PLAAVTLVLL REGAQKYFAG
                  SFYRGR*
              351
10
    ORF121ng-1 and ORF121-1 show 97.5% identity in 356 aa overlap:
                              10
                                       20
                                                30
                                                         40
                                                                  50
                                                                           60
         orf121-1.pep
                       MYRRKGRGIKPWMGAGAAFAALVWLVFALGDTLTPFAVAAVLAYVLDPLVEWLOKKGLNR
                       orf121ng-1
                       MYRRKGRGIKPWMGAGAAFAALVWLVYALGDTLTPFAVAAVLAYVLDPLVEWLQKKGLNR
15
                              10
                                       20
                                                30
                                                         40
                                                                  50
                              70
                                       80
                                                90
                                                        100
                                                                 110
         orf121-1.pep
                       ASASMSVMVFSLILLLALLLIIVPMLVGQFNNLASRLPOLIGFMONTLLPWLKNTIGGYV
                       20
         orf121ng-1
                       ASASMSVMVFSLILLLALLLIIVPMLVGQFNNLASRLPQLIGFMQNTLLPWLKNTIGGYV
                              70
                                       80
                                                90
                                                        100
                                                                 110
                             130
                                      140
                                               150
                                                        160
                                                                 170
                       EIDQASIIAWLQAHTGELSNALKAWFPVLMRQGGNIVSSIGNLLLLPLLLYYFLLDWORW
25
                       orf121ng-1
                       EIDQASIIAWFQAHTGELSNALKAWFPVLMKQGGNIVSSIGNLLLPPLLLYYFLLDWQRW
                             130
                                      140
                                               150
                                                        160
                                                                 170
                             190
                                      200
                                               210
                                                        220
                                                                 230
                                                                          240
30
         orf121-1.pep
                       SCGIAKLVPRRFAGAYTRITGNLNEVLGEFLRGQLLVMLIMGLVYGLGLVLVGLDSGFAI
                       orf121ng-1
                       SCGIAKLVPRRFAGAYTRITGNLNEVLGEFLRGQLLVMLIMGLVYGLGLMLVGLDSGFAI
                             190
                                      200
                                               210
                                                        220
                                                                 230
                                                                          240
35
                             250
                                      260
                                               270
                                                        280
                                                                 290
                       GMLAGILVFVPYLGAFTGLLLATVAALLQFGSWNGILSVWAVFAVGOFLESFFITPKIVG
          orf121-1.pep
                       GMVAGILVFVPYLGAFTGLLLATVAALLQFGSWNGILAVWAVFAVGQFLESFFITPKIVG
          orf121ng-1
                             250
                                      260
                                               270
                                                        280
                                                                 290
40
                             310
                                      320
                                               330
                                                        340
          orf121-1.pep
                       DRIGLSPFWVIFSLMAFGQLMGFVGMLAGLPLAAVTLVLLREGVQKYFAGSFYRGRX
                       orf121ng-1
                       DRIGLSPFWVIFSLMAFGELMGFVGMLAGLPLAAVTLVLLREGAQKYFAGSFYRGRX
45
                             310
                                      320
                                               330
                                                        340
     In addition, ORF121ng-1 shows homology to a permease from H.influenzae:
           sp|P43969|PERM HAEIN PUTATIVE PERMEASE PERM HOMOLOG Length = 349
           Score = 69.9 bits (168), Expect = 2e-11
           Identities = 67/317 (21%), Positives = 120/317 (37%), Gaps = 7/317 (2%)
50
          Query: 26
                   VYALGDTLTPFAVAAVLAYVLDPLVEWL-QKKGLNRASASMSVMVFSXXXXXXXXXXVP 84
                    +Y GD + P +A VL+Y+L+ + +L Q
                                                  R A++ +
                    IYFFGDLIAPLLIALVLSYLLEIPINFLNQYLKCPRMLATILIFGSFIGLAAVFFLVLVP 91
          Sbjct: 32
55
                   MLVGQFNNLASRLPQLIGFMQNTLLPWLKNTIGGYVE-IDQASIIAWFQAHTGELSNALK 143
          Ouerv: 85
                    ML Q +L S LP +
                                      N
                                           WL N
                                                  Y E ID + + + F +
          Sbjct: 92 MLWNQTISLLSDLPAMF----NKSNEWLLNLPKNYPELIDYSMVDSIFNSVREKILGFGE 147
          Query: 144 AWFPVLMKQGGNIVSSIGNXXXXXXXXXXXXXDWQRWSCGIAKLVPRRFAGAYTRITGNL 203
60
                                                       G+++ +P+
          Sbjct: 148 SAVKLSLASIMNLVSLGIYAFLVPLMMFFMLKDKSELLQGVSRFLPKNRNLAFXRWK-EM 206
          Query: 204 NEVLGEFLRGQXXXXXXXXXXXXXXXXXXXXDSGFAIGMVAGILVFVPYXXXXXXXXXXX 263
                                                        G+ V VPY
65
          Sbjct: 207 QQQISNYIHGKLLEILIVTLITYIIFLIFGLNYPLLLAFAVGLSVLVPYIGAVIVTIPVA 266
```

```
Query: 264 XXXXXQFGSWNGILAVWAVFAVGQFLESFFITPKIVGDRIGLSPFWVIFSLMAFGELMGF 323
QFG + FAV Q L+ + P + + L P + I S++ FG L GF
Sbjct: 267 LVALFQFGISPTFWYIIIAFAVSQLLDGNLLVPYLFSEAVNLHPLIIIISVLIFGGLWGF 326
Query: 324 VGMLAGLPLAAVTLVLL 340
G+ +PLA + ++
Sbjct: 327 WGVFFAIPLATLVKAVI 343
```

Based on this analysis, including the presence of a putative leader sequence and transmembrane domains in the two proteins, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 94

5

The following partial DNA sequence was identified in N.meningitidis <SEQ ID 791>:

```
..ACTGCTTTTT CGGCGGCGCT GCGCTTGAGT CCATCATGAC TCGTCATATT
                       TTTGTCCTTT GGGAAACCGT ATCAACAAAC AGCCGCCATC TTAACATTTT
15
                 51
                       TTTGCACGTC CTGCCCGCCG CGTTCAAATG CGTACCAGCA ATACCGCCGC
                101
                       CTGCGCCTCT ATGCCTTCCA TCCGCCCGAG ATAGCCGAGT TTTTCGTTGG
                151
                       TTTTGCCTTT GATGTTGACG CACGAAATGT CTATGCCCAA ATCGGCGGCG
                201
                       ATGTTGGCAC GCATTTGCGG AATGTGCGGC GCGAGTGTGG GTTTCTGTGC
                251
                       AATCACGGTC GTATCGACAT TGACCGCCTG CCAACCCTGC GCCTGAACGC
20
                301
                       TTTGATACGC CGCACGCAAA AGGACGCGGC TGTCCGCATC TTTGAACTCT
                351
                       GCGGCGGTGT CGGGGAAATG GCTGCCGATA TCGCCCAAAC CTGCCGCACC
                401
                       GAGCAGCGCG TCGGTAACGG CGTGCAGCAG CGCATCGGCA TCGGAGTGTC
                451
                       CGAGCAGCCC TTTTTCAAAT GGGATTTCAA CTCCGCCAAG TATCAG..
                501
```

25 This corresponds to the amino acid sequence <SEQ ID 792; ORF122>:

```
1 ..TAFSAALRLS PSXLVIFLSF GKPYQQTAAI LTFFCTSCPP RSNAYQQYRR
51 LRLYAFHPPE IAEFFVGFAF DVDARNVYAQ IGGDVGTHLR NVRRECGFLC
101 NHGRIDIDRL PTLRLNALIR RTQKDAAVRI FELCGGVGEM AADIAQTCRT
151 EQRVGNGVQQ RIGIGVSEQP FFKWDFNSAK YQ..
```

30 Further work revealed the complete nucleotide sequence <SEQ ID 793>:

```
1 ATATCGTACT GGGCAAGCAG TTCGCCGGAT TTTTTGGAAG TAGATACCGC
                    GCCTTTGATT TTTTTGCCGC TCTTACCCAA GGCTTCGATG AAAAAGTTGA
                51
                    TGGTCGAGCC GGTACCGATG CCGATATATT CATTTTCGGG TACGAATTCG
               101
                    ACTGCTTTTT CGGCGGCGAT GCGCTTGAGT TCGTCTTGTG TCGTCATATT
               151
                    TTTGTCCTTT GGGAAACCGT ATCAACAAAC AGCCGCCATC TTAACATTTT
35
               201
                    TTTGCACGTC CTGCCCGCCG CGTTCAAATG CGTACCAGCA ATACCGCCGC
               251
                    CTGCGCCTCT ATGCCTTCCA TCCGCCCGAG ATAGCCGAGT TTTTCGTTGG
               301
                351 TTTTGCCTTT GATGTTGACG CACGAAATGT CTATGCCCAA ATCGGCGGCG
                401 ATGTTGGCAC GCATTTGCGG AATGTGCGGC GCGAGTTTGG GTTTCTGTGC
                    AATCACGGTC GTATCGACAT TGACCGCCTG CCAACCCTGC GCCTGAACGC
40
                451
                    TTTGATACGC CGCACGCAAA AGGACGCGGC TGTCCGCATC TTTGAACTCT
                501
                    GCGGCGGTGT CGGGGAAATG GCTGCCGATA TCGCCCAAAC CTGCCGCACC
                551
                    GAGCAGCGCG TCGGTAACGG CGTGCAGCAG CGCATCGGCA TCGGAGTGTC
                601
                     CGAGCAGCCC TTTTTCAAAT GGGATTTCAA CTCCGCCAAG TATCAGCTTT
                651
                     CTGCCTTCGG TCAGTTGGTG GACATCGTAG CCCTGTCCGA TACGGATGTT
45
                701
                751 CGTCATCGTT TGTGTTCCTG A
```

This corresponds to the amino acid sequence <SEQ ID 794; ORF122-1>:

```
1 ISYWASSSPD FLEVDTAPLI FLPLLPKASM KKLMVEPVPM PIYSFSGTNS
51 TAFSAAMRLS SSCVVIFLSF GKPYQQTAAI LTFFCTSCPP RSNAYQQYRR
101 LRIYAFHPPE IAEFFVGFAF DVDARNVYAQ IGGDVGTHLR NVRREFGFLC
151 NHGRIDIDRL PTLRLNALIR RTQKDAAVRI FELCGGVGEM AADIAQTCRT
201 EQRVGNGVQQ RIGIGVSEQP FFKWDFNSAK YQLSAFGQLV DIVALSDTDV
251 RHRLCS*
```

Computer analysis of this amino acid sequence gave the following results:

# Homology with a predicted ORF from N. meningitidis (strain A)

ORF122 shows 94.0% identity over a 182aa overlap with an ORF (ORF122a) from strain A of N. meningitidis:

					•		
	•				10	20	30
5	orf122.pep			TAFS:	AALRLSPSXLVIE		
	OZZZZZ POP				:      :		
	orf122a	FLPLLPKASMKKLM	/EPVPMPMYS	FSGTNSTAFS	AAMRISSSCVVTF	T.SEGKPYO	ОТААТ
		30		50	60	70	80
						. •	•
10		40	50	60	70	80	90
	orf122.pep	LTFFCTSCPPRSNA	<b>YQQYRRLRL</b> Y	AFHPPEIAEF	FVGFAFDVDARNV	YAOIGGDV	GTHLR
	•	1111 1111111	[]]]]]	111 111:11	ини пин	THEFT	11111
	orf122a	LTFFXTSCPPRSNP	YQQYRRLRL\	AFHAPEITEF:	FVGFAFXVDARNV	YAQIGGDV	GTHLR
		90	100	110	120	130	140
15							
	•	100	110	120	130	140	150
	orf122.pep	NVRRECGFLCNHGR:	IDIDRLPTLE	RLNALIRRTQK!	DAAVRIFELCGGV	GEMAADIA	QTCRT
		1:111 11111111			[[[]]]	11111111	11111
20	orf122a	NMRREFGFLCNHGR:	IDIDRLPTLF	RLNALIRRTQK	DAAVRIFELCGGV	GEMAADIA	.QTCRT
20	••	150	160	170	180	190	200
			•				
		160	170	180			
	orf122.pep	EQRVGNGVQQRIGIO	SVSEOPFFKV	idfnsakyo			
25		111111111111111	111111111	1111111			
25	orf122a	EQRVGNGVQQRIGIO					SX
		210	220	230	240	250	
	The complete length	ORF122a puele	otide ceau	ance CTO	TD 705> :		
	The complete length	i Oitt 122a nucit	oude sequ	cuce -pro	1937 18:		
•	1 200		2010 8801				
	1 ATA 51 GCC	ATCATATT GGGCAA	CAG TTCAC	CIGGAT TTTT	TGGAAG TAGATA	rccec	
30	101 TG	CTTTGATT TTTTTGG STCGAACC GGTACCG	COC TOTAL	ACCCAA GGCT	TCGATG AAAAAG	TTGA	
50	151 AC	FGCNTTTT CGGCGG	CAM CCCA	TGTATT CGTT	TTCGGG TACGAA	ATTCG	
	201 TT	rgentiti eggegg rgteettt eggaaa	CCT ATCA	CARGE TOGT	COTTGTG TCGTCA	TATT	
	251 TTI	NNACGTC CTGCCC	CCC CCTTC	ACAAAC AGCC	GCCATC TTAACA	TTTT	
	301 CT	GCGACTCT ATGCCT	CCG CGIIC	CCCAC ATTAL	CCCAGCA ATACCG	JCCGC	
35	351 TT	TTGCCTTT GANGTT	TACE CACE	ADDCT CTAM	CCCCAA AMOCCC	TTGG	
55	401 AT	GTTGGCAC GCATTT	CCC DATA	GCGGC GCGM	CTTTCC CTTTCC	CUCC	
	451 AA	CACGGTC GTATCG	ACAT TGAC	CCCCC CCAN	CCCACC CCCAC	ACCC	
	501 TT	TGATACGC CGCACG	CAAA AGGAC	CCCCC TCTC	CCCIGC GCCIGA	MCGC	
	551 GC	GCGGTGT CGGGGA	AATG GCTGC	CGATA TOGO	CCAAAC CTGCCG	CACC	
40	601 GA	GCAGCGCG TCGGTA	ACGG CGTGC	AGCAG CGCA	TOGGOA TOGGAG	TGTC	
	651 CG/	AGCAGCCC TTTTTC	AAAT GGGAT	TTCAA CTCC	GCCAAG TATCAG	,1010 ,1010	
	701 CT	SCCTTCGG TCAGTT	GTG GACAT	CGTAG CCCT	GTCCGA TACGGA	, итстт итстт	
	751 CG	CATCGTT TGTGTT	CCTG A				
	This encodes a prote	ein having amino	acid sequ	ence <seo< th=""><th>ID 796&gt;:</th><th></th><th></th></seo<>	ID 796>:		
		•	•	`			
45	1 IS	YWASSSLD FLEVDT	APLI FLPLI	PKASM KKLM	VEPVPM PMYSES	GTNS	
	51 TA	FSAAMRLS SSCVVI	FLSF GKPY	OTAAI LTFF	XTSCPP RSNPYC	MYRR	
	101 LR	LYAFHAPE ITEFFV	GFAF XVDA	RNVYAO IGGD	VGTHLR NMRREE	GFLC	
	151 NH	GRIDIDRL PTLRLN	ALIR RTOKI	DAAVRI FELC	GGVGEM AADTAC	TCRT	
	201 EQ	RVGNGVQQ RIGIGV	SEQP FFKWI	FNSAK YQLS	AFGOLV DIVALS	EDTDV	
50	251 RH	RLCS*		-			
	000000 1000				_		
	ORF122a and ORF	122-1 show 96.9°	% identity	in 256 aa ov	erlap:		
					_		
		10	20	30	40	50	60
	orf122a.pep	ISYWASSSLDFL	EVDTAPLIFI	LPLLPKASMKK	LMVEPVPMPMYSE		'SAAMRI.S
		111111111111111111111111111111111111111	111111111		11111111:11		1111111
55	orf122-1	ISYWASSSPDFL	EVDTAPLIF	PLLPKASMKK	LMVEPVPMPIYS	SGTNSTAF	SAAMRIS
		10	20	30	40	50	60
					. ==		•
	·	70	80	90	100	110	120
	orf122a.pep	SSCVVIFLSFGK	PYQQTAAIL1	FFXTSCPPRS	NPYQQYRRLRLY <i>I</i>	AFHAPEITE	FFVGFAF
60	•	11111111111	111111111		1 1111111111	111 111:1	HILLI
	orf122-1	SSCVVIFLSFGK	PYQQTAAIL	FFCTSCPPRS	NAYQQYRRLRLY	AFHPPEIAF	FFVGFAF
						. –	

-441-

			-441	.=				
		70	80	90	100	110	120	
5	orf122a.pep	130 XVDARNVYAQIGGDV(           DVDARNVYAQIGGDV( 130	11:11111	1111111111	1111111111	11111111111	111111	
10	orf122a.pep orf122-1	190 FELCGGVGEMAADIA(               FELCGGVGEMAADIA( 190	HHHHH	1111111111	111111111	ППППП	1111111	
15 20	orf122a.pep orf122-1	250 DIVALSDTDVRHRLC:            DIVALSDTDVRHRLC: 250	11		·			
20		redicted ORF from N 6% identity over a 1		· <del>·····</del>	predicted	ORF (ORI	F122ng) fro	om
	N.gonorrhoeae:							
25	orf122.pep	FLPLLPKASMKKLMVEP		 GTNSTAFSA	:      :   MRLSSSCVV		IIIII OTAAI 80	) ¦
30	orf122.pep orf122ng orf122.pep	LTFFCTSCPPRSNAYQQ'	 YRRLRLYAF	 HPPEIAEFFV	:     GFAFDIDAR	: :      NIDTQIGGDV	 GTHLR 140	,
35	orf122ng		:        DHLPTLRLN			11:111:1		
	orf122ng The complete lengt			   NSAKYQLSAE	-			
40	51 GC 101 tg	GTCGTACC GGGCAAGCA CTTTGATT TTTTTACCG GTCGAACC GGtaCCGAT TGCTTTTT CGGCGGCGA	C TTTTGCC G CCGATGT	CAA GGCTTC	GATG AAGAI	ATTGa ATTCG		
45	201 TT 251 TT 301 ct 351 TT 401 AT	TAtcettt gGGAAacce TGCACGtc ctggccgce gcgcctCT AtgcCTTCC TTGCCTTT GATatTGAC GTTGGCAC GCATTTGCG	t atcaAca g cgttcaA A TCCGCCC G CACGAAA G AATGTGC	AAC agccgo ATC cgtaco CGAG ATAGCO ATAT CGatao CGGT GCGAGT	catC TTAA( aGca atac GAGT TTTT( CCAa atcg TTGG GTTT	CATTTT CGCCGC CGTTGG JCGGCG CTGTGC		
50	501 TT 551 GC 601 GA 651 CG	TCACGGTC GTATCGACA TGATACGC CGCACGCAA GGCGGTGT CGGGAAAAT GCAGCGCG tcggtaaCG AGCAGCCC TTTTTCAAA	A AGGACGO G GCTGCCG G CGTGCAG T GGGATTT	CGGC TGTCCC CATG TCGCCC CCAG CGCGTC CCAA CTCCGC	CATC TTTGA CAAAC CTGCG GGCA TCCGA CCAAG TATCA	AACTCT CGCACC AATGCC AGCTTT		
55		GCCTTCGG TCAATTGGT TCATCGTT TGTGTTCCT		STAG CCCTGT	CCGA TACG	SATATT		
	This encodes a prot	tein having amino ac	id sequen	ce <seq ii<="" th=""><th>798&gt;:</th><th></th><th></th><th></th></seq>	798>:			
	1 MS	YRASSSPD FLEVETAPL	I FLPLLPK	KASM KKLMVI	PVPM PMYS	esgens		

	1	MSYRASSSPD	FLEVETAPLI	FLPLLPKASM	KKLMVEPVPM	PMYSFSGTNS	
	51	TAFSAAMRLS	SSCVVIFLSF	GKPYQQTAAI	LTFFCTSWPP	RSNPYQQYRR	
	101	LRLYAFHPPE	IAEFFVGFAF	DIDARNIDTQ	IGGDVGTHLR	NVRCEFGFLC	
60	151	NHGRIDIDHL	PTLRLNALIR	RTQKDAAVRI	FELCGGVGKM	AADVAQTCRT	
	201	EQRVGNGVQQ	RVGIRMPEQP	FFKWDFNSAK	YQLSAFGQLV	DIVALSDTDI	
	251	RHRLCS*					

## ORF122ng and ORF122-1 show 92.6% identity in 256 aa overlap:

		10	20	30	40	· 50	60
	orf122-1.pep	ISYWASSSPDFLE	VDTAPLIFLPL	LPKASMKKLI	<i>I</i> VEPVPMPIYS	FSGTNSTAFS	AAMRLS
_		: 11	<b>!:       </b>	11111111	[11]][1]:[1	HHHHHH	HHHI
5	orf122ng	MSYRASSSPDFLE			<i>i</i> vepvpmpmys	FSGTNSTAFS	AAMRLS
	,	10	20	30	40	50	60
		70	80	90	100	110	120
	orf122-1.pep	SSCVVIFLSFGKP					
10					1111111111		HILLI
	orf122ng	SSCVVIFLSFGKP	YOOTAAILTEF	CTSWPPRSNI	PYOOYRRLRLY	AFHPPETAER	
	•	70	80	90	100	110	120
		130	140	150	160	170	180
15	orf122-1.pep	DVDARNVYAQIGG	DVGTHLRNVRRI	EFGFLCNHG	RIDIDRLPTLR	LNALIRRTON	CDAAVRI
		1:1111: :1111			11111:11111		ніш
	orf122ng	DIDARNIDTQIGG		EFGFLCNHGI	RIDIDHLPTLR	LNALIRRTOK	DAAVRI
		130	140	150	160	170	180
20						•	
20		190	200	210	220	230	240
	orf122-1.pep	FELCGGVGEMAAD		GNGVQQRIG:	IGVSEQPFFKW	<i>I</i> DFNSAKYQLS	SAFGQLV
	C1 00 .			1111111:1	, , ,,,,,,,		
	orf122ng	FELCGGVGKMAAD	VAQTCRTEQRV				
25		190	200	210	220	230	240
23	4.	250					
•	orf122-1.pep	DIVALSDTDVRHR	T CCV		•		
	orriss-r.beb						
	orf122ng	DIVALSDTDIRHR					
30	OLLIZZING	250	DCSA				
-		250					

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 95

The following partial DNA sequence was identified in N.meningitidis <SEQ ID 799>:

```
1 ...GCCGGCGCGA GTGCGAACAA CATTTCCGCG CGTTTTGCGG AAACACCCCGT
51 CGCTGTCAGC GTTACCCTGA TCGGCACGGT ACTTGCCGTC ATGCTGCCCG
101 TTACCGAATA TGAAAACTTC CTGCTGCTTA TCGGCTCGGT ATTTGCGCCG
151 ATGGGGCGGA TTTTGATTGC CGACTTTTTC GTCTTGAAAC GGCGTGA
```

This corresponds to the amino acid sequence <SEQ ID 800; ORF125>:

1 ... AGASANNISA RFAETPVAVS VTLIGTVLAV MLPVTEYENF LLLIGSVFAP
51 MGGFDCRLFR LETA\*

### Further work revealed the complete nucleotide sequence <SEQ ID 801>:

	1	ATGTCGGGCA	ATGCCTCCTC	TCCTTCATCT	TCCTCCGCCA	TCGGGCTGAT
	51	TTGGTTCGGC	GCGGCGGTAT	CGATTGCCGA	AATCAGCACG	GGTACGCTGC
45	101	TTGCGCCTTT	GGGCTGGCAG	CGCGGTCTGG	CGGCTCTACT	TTTGGGTCAT
	151	GCCGTCGGCG	GCGCGCTGTT	TTTTGCGGCG	GCGTATATCG	GCGCACTGAC
	201	CGGACGCAGC	TCGATGGAAA	GCGTGCGCCT	GTCGTTCGGC	AAACGCGGTT
	251	CAGTGCTGTT	TTCCGTGGCG	AATATGCTGC	AACTGGCCGG	CTGGACGGCG
	301	GTGATGATTT	ACGCCGGCGC	AACGGTCAGC	TCCGCTTTGG	GCAAAGTGTT
50	351	GTGGGACGGC	GAATCTTTTG	TCTGGTGGGC	ATTGGCAAAC	GGCGCGCTGA
	401	TTGTGCTGTG	GCTGGTTTTC	GGCGCACGCA	AAACAGGCGG	GCTGAAAACC
	451	GTTTCGATGC	TGCTGATGCT	GTTGGCGGTT	CTGTGGCTGA	GTGCCGAAGT
	501	CTTTTCCACG	GCAGGCAGCA	CCGCCGCACA	GGTTTCAGAC	GGCATGAGTT
	551		AGTCGAGCTG			
55	601	CTTGCCGCCG	ACTACACGCG	CCACGCGCGC	CGCCCGTTTG	CGGCAACCCT
	651		CTCGCCTACA			
	701		GGCGTTGTTC			
•	751					TCCTCTCCAC

	801	CGTTACCACA	ACGTTTCTCG	ATGCCTATTC	CGCCGGCGCG	AGTGCGAACA
	851	ACATTTCCGC	GCGTTTTGCG	GAAACACCCG	TCGCTGTCGG	CGTTACCCTG
	901	ATCGGCACGG	TACTTGCCGT	CATGCTGCCC	GTTACCGAAT	ATGAAAACTT
	951	CCTGCTGCTT	ATCGGCTCGG	TATTTGCGCC	GATGGCGGCG	GTTTTGATTG
5	1001	CCGACTTTTT	CGTCTTGAAA	CGGCGTGAGG	AGATTGAAGG	CTTTGACTTT
	1051	GCCGGACTGG	TTCTGTGGCT	TGCGGGCTTC	ATCCTCTACC	GCTTCCTGCT
	1101	CTCGTCCGGC	TGGGAAAGCA	GCATCGGTCT	GACCGCCCCC	GTAATGTCTG
	1151	CCGTTGCCAT	TGCCACCGTA	TCGGTACGCC	TTTTCTTTAA	AAAAACCCAA
	1201	TCTTTACAAA	GGAACCCGTC	ATGA		

10 This corresponds to the amino acid sequence <SEQ ID 802; ORF125-1>:

```
MSGNASSPSS SSAIGLIWFG AAVSIAEIST GTLLAPLGWQ RGLAALLLGH
                51
                     AVGGALFFAA AYIGALTGRS SMESVRLSFG KRGSVLFSVA NMLQLAGWTA
                    VMIYAGATVS SALGKVLWDG ESFVWWALAN GALIVLWLVF GARKTGGLKT
                101
                    VSMLLMLLAV LWLSAEVFST AGSTAAQVSD GMSFGTAVEL SAVMPLSWLP
                151
                201 LAADYTRHAR RPFAATLTAT LAYTLTGCWM YALGLAAALF TGETDVAKIL
15
                251
                    LGAGLGAAGI LAVVLSTVTT TFLDAYSAGA SANNISARFA ETPVAVGVTL
                301
                     IGTVLAVMLP VTEYENFLLL IGSVFAPMAA VLIADFFVLK RREEIEGFDF
                351
                     AGLVLWLAGF ILYRFLLSSG WESSIGLTAP VMSAVAIATV SVRLFFKKTQ
                401
                     SLQRNPS*
```

20 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from N. meningitidis (strain A)

ORF125 shows 76.5% identity over a 51aa overlap with an ORF (ORF125a) from strain A of N. meningitidis:

```
10
                                                                     20
25
          orf125.pep
                                                    AGASANNISARFAETPVAVSVTLIGTVLAV
                                                    11:1111111:::1 1:11:1::::11:11
          orf125a
                       KILLGAGLGAAGILAVVLSTVTTTFLDAYSAGVSANNISAKLSEIPIAVAVAVVGTLLAV
                                260
                                          270
                                                    280
                                                             290
                                                                       300
30
                               40
                                        50
                                                  60
                       MLPVTEYENFLLLIGSVFAPMGGFDCRLFRLETAX
          orf125.pep
                       orf125a
                       LLPVTEYENFLLLIGSVFAPMAAVLIADFFVLKRREEIEG
                       310
                                320
                                          330
```

35 The ORF125a partial nucleotide sequence <SEQ ID 803> is:

	1	ATGTCGGGCA	ATGCCTCCTC	TCNTTCATCT	TCCGCCGCCA	TCGGGCTGAT
	51	TTGGTTCGGC	GCGGCGGTAT	CGATTGCCGA	AATCAGCACG	GGTACACTGC
	101	TTGCGCCTTT	GGGCTGGCAG	CGCGGTCTGG	CNGCTCTGCT	TTTGGGTCAT
	151	GCCGTCGGCG	GCGCGCTGTT	TTTTGCGGCG	GCGTATATCG	GCGCACTGAC
40	201	CGGACNCANC	TCGATGGAAA	GCGTGCGCCT	GTCGTTCGGC	AAACGCGGTT
	251	CAGTGCTGTT	TTCCGTGGCG	AATATGCTGC	AACTGGCCGG	CTGGACGGCG
	301	GTGATGATTT	ACGCCGGCGC	AACGGTCAGC	TCCGCTTTGG	GCAAAGTGTT
	351	GTGGGACGGC	GAATCTTTTG	TCTGGTGGGC	ATTGGCAAAC	GGCGCGCTGA
	401	TTGTGCTGTG	GCTGGTTTTC	GGCGCACGCA	AAACAGGCGG	GCTGAAAACC
45	451	GTTTCGATGC	TGCTGATGCT	GTTGGCGGTT	CTGTGGCTGA	GTGCCGAANT
	501		GCAGGCAGCA	CCGCCGCANN	GGTNNCAGAC	GGCATGAGTT
	551	TCGGAACGGC	AGTCGAGCTG	TCCGCCGTNA	TGCCGCTTTC	TTGGCTGCCG
	601	CTGGCCGCCG	ACTACACGCG	CCACGCGCGC	CGCCCGTTTG	CGGCAACCCT
	651	GACGGCAACG	CTCGCCTACA	CGCTGACCGG	CTGCTGGATG	TATGCCTTGG
50	701	GTTTGGCAGC	GGCGTTGTTC	ACCGGAGAAA	CCGACGTGGC	AAAAATCCTG
	751	CTGGGCGCAG	GTTTGGGTGC	GGCAGGCATT	TTGGCGGTCG	TCCTGTCGAC
	801	CGTTACCACC	ACTTTTCTCG	ATGCNTACTC	CGCCGGCGTA	AGTGCCAACA
	851	ATATTTCCGC	CAAACTTTCG	GAAATACCNA	TCGCCGTTGC	CGTCGCCGTT
	901	GTCGGCACAC	TGCTTGCCGT	CCTCCTGCCC	GTTACCGAAT	ATGAAAACTT
55	951	CCTGCTGCTT	ATCGGCTCGG	TATTTGCGCC	GATGGCGGCG	GTTTTGATTG
	1001	CCGACTTTTT	CGTCTTGAAA	CGGCGTGAGG	AGATTGAAGG	C

This encodes a protein having the partial amino acid sequence <SEQ ID 804>:

<sup>1</sup> MSGNASSXSS SAAIGLIWFG AAVSIAEIST GTLLAPLGWQ RGLAALLLGH

<sup>51</sup> AVGGALFFAA AYIGALTGXX SMESVRLSFG KRGSVLFSVA NMLQLAGWTA

	101 151	VMIYAGATVS			
		VSMLLMLLAV LAADYTRHAR			
	251	LGAGLGAAGI			 
5	301	VGTLLAVLLP		_	
	 	DD1051	0.4.504.11		 •

#### ORF125a and ORF125-1 show 94.5% identity in 347 aa overlap:

	, <b>•</b>	10	20	30	40	50	60
10	orf125a.pep	MSGNASSXSSSAA					
10	cror 1	111111111111111111111111111111111111111					
	orf125-1	MSGNASSPSSSSA					
		10	20	30	40	50	60
		70	80	90	. 100	110	120
15	orf125a.pep	AYIGALTGXXSME	SVRLSFGKRGS	SVLFSVANMLQ	LAGWTAVMIY	AGATVSSAL	GKVLWDG
			1111111111		1111111111		
	orf125-1	AYIGALTGRSSME		SVLFSVANMLQ	LAGWTAVMIY	AGATVSSAL	GKVLWDG
		70	. 80	90	100	110	120
20							
20		130	140	150	160	170	180
	orf125a.pep	ESFVWWALANGAL	IVLWLVFGAR	KTGGLKTVSML	LMLLAVLWLS	AEXFSTAGS	TAAXVXD
		POTENTIAL ANGAY	7777 777 777 777			11 111111	
	orf125-1	ESFVWWALANGAL 130	140	TGGLKTVSML	LMLLAVLWLS 160		
25		130	140	150	100	170	180
23		190	200	210	220	230	240
	orf125a.pep	GMSFGTAVELSAV					
						111111111	
	orf125-1	GMSFGTAVELSAV	MPLSWLPLAAI	DYTRHARRPFA	ATLTATLAYI	LTGCWMYAL	GLAAALF
30		190	200	210	220	230	240
					•		
		250	260	270	280	290	300
	orf125a.pep	TGETDVAKILLGA	GLGAAGILAV	VLSTVTTTFLC	AYSAGVSANN	ISAKLSEIP	VAVAV
25							:  : ::
35	orf125-1	TGETDVAKILLGA					
		250	260	270	280	290	300
		310	320	330	340		
	orf125a.pep	VGTLLAVLLPVTE				TEC	
40	Offizsa.pep	:11:111:1111	I I I I I I I I I I I I I I I I I I I	ALWENWAATTW	DEEAPVEC	111	
.0	orf125-1	IGTVLAVMLPVTE	YENFI.I.T.CS	IIIIIIIIIIII VEADMAAVT.TA		III Tegeneact	VIMIACE
		310	320	330	340	350	360
					<b>4.0</b>	550	300

#### Homology with a predicted ORF from N.gonorrhoeae

ORF125 shows 86.2% identity over a 65aa overlap with a predicted ORF (ORF125ng) from N.gonorrhoeae:

An ORF125ng nucleotide sequence <SEQ ID 805> was predicted to encode a protein having amino

55 acid sequence <SEQ ID 806>:

```
1 MSGNASSPSS SAAIGLVWFG AAVSIAEIST GTLLAPLGWQ RGLAALLIGH
51 AVGGALFFAA AYIGALTGRS SMESVRLSFG KCGSVLFSVA NMLQLAGWTA
101 VMIYVGATVS SALGKVLWDG ESFVWWALAN GALIVLWLVF GARRTGGLKT
151 VSMLLMLLAV LWLSVEVFAS SGTNAAPAVS DGMTFGTAVE LSAVMPLSWL
201 PLAADYTRQA RRPFAATLTA TLAYTLTGCW MYALGLAAAL FTGETDVAKI
251 LLGAGLGITG ILAVVLSTVT TTFLDTYSAG ASANNISARF AEIPVAVGVT
301 LIRTVLAVML PVTEYKNFLL LIRSVFGPMA GGFDCRLFCL KTA*
```

1 ATGTCGGGCA ATGCCTCCTC TCCTTCATCT TCCGCCGCCA TCGGGCTGGT

Further work revealed the following gonococcal DNA sequence <SEQ ID 807>:

					TCCGCCGCCA		
					AATCAGCACG		
_	101 TCG	CCCCCTT	GGGCTGGCAG	CGCGGTCTGG	CGGCCCTGCT	TTTGGGTCAT	
5	151 GCC	GTCGGCG	GCGCGCTGTT	TTTTGCGGCG	GCGTATATCG	GCGCACTGAC	
	201 CGG	ACGCAGC	TCGATGGAAA	GTGTGCGCCT	GTCGTTCGGC	AAATGCGGTT	
					AACTGGCCGG		
					TCCGCTTTGG		
					ATTGGCAAAC		
10							
10					GAACGGGCGG		
					TTGTGGTTGA		
					CGCCGTTTCA		
	551 CCT	TCGGAAC	GGCAGTCGAA	CTGTCCGCCG	TCATGCCGCT	TTCCTGGCTG	
					CGCCGCCCGT		
15	651 CCT	GACGGCA	ACGCTCGCCT	ATACGCTGAC	GGGCTGCTGG	ATGTATGCCT	
					AAACCGACGT		
					ATTCTGGCAG		
					TTCCGCCGGC		
20		ACATTTC	CGCGCGTTTT	GCGGAAATAC	CCGTCGCTGT	CGGCGTTACC	
20	901 CTG	ATCGGCA	CGGTGCTTGC	CGTCATGCTG	CCCGTTACCG	AATATAAAAA	
					GCCGATGGCG		
		CCGACTT	TTTCGTCTTA	AAACGGCGTG	AGGAGATTGA	AGGCTTTGAC	
	1051 TTT:	GCCGGAC	TGGTTCTGTG	GCTGGCAGGC	TTCATCCTCT	ACCGCTTCCT	
	1101 GCT	CTCGTCC	GGTTGGGAAA	GCAGCATCGG	TCTGACCGCC	CCCGTAATGT	
25					GCCTTTTCTT		
			AAAGGAACCC		0001111011	ITTTTTTTTCC	
	1201 CAA	ICITIAC	AAAOOAACCC	GICAIGA			
	This corresponds to	the amin	o acid ceane	nce <ceo ii<="" th=""><th>7 606. UDE</th><th>125ng 1&gt;.</th><th>,</th></ceo>	7 606. UDE	125ng 1>.	,
	Tims corresponds to	me aimi	o acid seque	Tre ZEG II	J 606, OKI	12311g-1/.	
	1 MSG	NASSPSS	SAAIGLVWFG	AAVSIAEIST	GTLLAPLGWQ	RGLAALLLGH	
	51 AVG	GALFFAA	AYIGALTGRS	SMESVRLSFG	KCGSVLFSVA	NMLOLAGWTA	
30	101 VMI	YVGATVS	SALGKVLWDG	ESFVWWALAN	GALIVLWLVF	GARRTGGLKT	
	151 VSM	T.LMT.T.AV	LWLSVEVEAS	SCTNAAPAVS	DGMTFGTAVE	I.SAVMPT.SWI.	
	201 PLĀ	ΔΟΥΤΡΟΔ	PRPEAATITA	TIAVTITCOW	MYALGLAAAL	EACEADINE TOWN	
					ASANNISARF		
		TOTOTIO.		TITLUTIONG	<b>WOWNIATOWKE</b>	WETLAWAGAT	
				TTCCUTTTDATA	ALLE TABLET	WD DED TRAFF	
25	301 <u>LIG</u>	TVLAVML	PVTEYKNFLL		AVLIADFFVL		
35	301 <u>LIG</u> 351 FAG	TVLAVML LVLWLAG	PVTEYKNFLL		AVLIADFFVL PVMSAVAIAT		
35	301 <u>LIG</u> 351 FAG	TVLAVML	PVTEYKNFLL				
35	301 <u>LIG</u> 351 <u>FAG</u> 401 QSL	TVLAVML LVLWLAG QRNPS*	PVTEYKNFLL FILYRFLLSS	GWESSIGLTA	PVMSAVAIAT	VSVRLFFKKT	
35	301 <u>LIG</u> 351 FAG	TVLAVML LVLWLAG QRNPS*	PVTEYKNFLL FILYRFLLSS	GWESSIGLTA	PVMSAVAIAT	VSVRLFFKKT	
35	301 <u>LIG</u> 351 <u>FAG</u> 401 QSL	TVLAVML LVLWLAG QRNPS*	PVTEYKNFLL FILYRFLLSS	GWESSIGLTA	PVMSAVAIAT	VSVRLFFKKT	
35	301 <u>LIG</u> 351 <u>FAG</u> 401 QSL	TVLAVML LVLWLAG QRNPS*	PVTEYKNFLL FILYRFLLSS show 95.1%	GWESSIGLTA  6 identity in	PVMSAVAIAT 408 aa overl	<u>vsvrlff</u> kkt ap:	
35	301 <u>LIG</u> 351 <u>FAG</u> 401 QSL ORF125ng-1 and O	TVLAVML LVLWLAG QRNPS* RF125-1	PVTEYKNFLL FILYRFLLSS show 95.1%	GWESSIGLTA  6 identity in	408 aa overl	VSVRLFFKKT  ap: 40 5	0 60
	301 <u>LIG</u> 351 <u>FAG</u> 401 QSL	TVLAVML LVLWLAG QRNPS* RF125-1 MSGNAS	PVTEYKNFLL FILYRFLLSS show 95.1%	GWESSIGLTA  6 identity in  20 IWFGAAVSIAE	408 aa overl	VSVRLFFKKT  ap:  40 5  WQRGLAALLLG	0 60 HAVGGALFFAA
35 40	301 <u>LIG</u> 351 <u>FAG</u> 401 QSL  ORF125ng-1 and O	TVLAVML LVLWLAG QRNPS* RF125-1 MSGNAS	PVTEYKNFLL FILYRFLLSS show 95.1% 10 SSPSSSSAIGL	GWESSIGLTA  6 identity in  20 IWFGAAVSIAE:	408 aa overl	VSVRLFFKKT  iap:  40 5  WORGLAALLG	0 60 HAVGGALFFAA
	301 <u>LIG</u> 351 <u>FAG</u> 401 QSL ORF125ng-1 and O	TVLAVML LVLWLAG QRNPS* RF125-1 MSGNAS	PVTEYKNFLL FILYRFLLSS  show 95.19  10 SSPSSSSAIGLE	GWESSIGLTA  6 identity in  20 IWFGAAVSIAE:	408 aa overl 30 istgtllaplgi	VSVRLFFKKT  ap:  40 5  WQRGLAALLG	0 60 HAVGGALFFAA 
	301 <u>LIG</u> 351 <u>FAG</u> 401 QSL  ORF125ng-1 and O	TVLAVML LVLWLAG QRNPS* RF125-1 MSGNAS	PVTEYKNFLL FILYRFLLSS show 95.1% 10 SSPSSSSAIGL	GWESSIGLTA  6 identity in  20 IWFGAAVSIAE:	408 aa overl 30 istgtllaplgi	VSVRLFFKKT  iap:  40 5  WORGLAALLG	0 60 HAVGGALFFAA 
	301 <u>LIG</u> 351 <u>FAG</u> 401 QSL  ORF125ng-1 and O	TVLAVML LVLWLAG QRNPS* RF125-1 MSGNAS	PVTEYKNFLL FILYRFLLSS  show 95.19  10 SSPSSSSAIGLE	GWESSIGLTA  6 identity in  20 IWFGAAVSIAE:	408 aa overl 30 istgtllaplgi	VSVRLFFKKT  ap:  40 5  WQRGLAALLG	0 60 HAVGGALFFAA 
40	301 <u>LIG</u> 351 <u>FAG</u> 401 QSL  ORF125ng-1 and O	TVLAVML LVLWLAG QRNPS* RF125-1 MSGNAS	PVTEYKNFLL FILYRFLLSS  show 95.19  10 SSPSSSSAIGLE	GWESSIGLTA  6 identity in  20 IWFGAAVSIAE:	408 aa overl 30 (STGTLLAPLGE) ISTGTLLAPLGE ISTGTLLAPLGE 30 (STGTLLAPLGE) 30 (STGTLLAPLGE)	VSVRLFFKKT  ap:  40 5  WQRGLAALLG	0 60 HAVGGALFFAA            HAVGGALFFAA 0 60
	301 <u>LIG</u> 351 <u>FAG</u> 401 QSL  ORF125ng-1 and O	TVLAVML LVLWLAG QRNPS* RF125-1 MSGNAS IIIII MSGNAS	PVTEYKNFLL FILYRFLLSS  show 95.1%  10 SSPSSSSAIGLE	GWESSIGLTA  6 identity in  20 IWFGAAVSIAE: :         :WFGAAVSIAE: 20  80	408 aa overl 30 ISTGTLLAPLGI ISTGTLLAPLGI 30 90	VSVRLFFKKT  lap:  40 5 WQRGLAALLLG             WQRGLAALLLG 40 5	0 60 HAVGGALFFAA           HAVGGALFFAA 0 60
40	301 <u>LIG</u> 351 <u>FAG</u> 401 QSL  ORF125ng-1 and Of  orf125-1.pep  orf125ng-1	TVLAVML LVLWLAG QRNPS* RF125-1 MSGNAS             MSGNAS	PVTEYKNFLL FILYRFLLSS  show 95.19  10 SSPSSSSAIGL:               SSPSSSAAIGLY 10  70 LTGRSSMESVR	GWESSIGLTA  6 identity in  20 IWFGAAVSIAE:          VWFGAAVSIAE: 20  80 LSFGKRGSVLF:	408 aa overl 30 ISTGTLLAPLGI ISTGTLLAPLGI 30 90 10 SVANMLQLAGW	VSVRLFFKKT lap: 40 5 WQRGLAALLLG IIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	0 60 HAVGGALFFAA           HAVGGALFFAA 0 60 0 120 SSALGKVLWDG
40	301 <u>LIG</u> 351 <u>FAG</u> 401 QSL  ORF125ng-1 and Of  orf125-1.pep  orf125ng-1	TVLAVML LVLWLAG QRNPS* RF125-1 MSGNAS             MSGNAS	PVTEYKNFLL FILYRFLLSS  show 95.19  10 SSPSSSSAIGL:              SSPSSSAAIGLY 10  70 LTGRSSMESVR:	GWESSIGLTA  6 identity in  20 IWFGAAVSIAE:          VWFGAAVSIAE: 20  80 LSFGKRGSVLF:	PVMSAVAIAT  408 aa overl  30 ISTGTLLAPLGI ISTGTLLAPLGI 30  90 ISVANMLQLAGW	VSVRLFFKKT    ap:   40	0 60 HAVGGALFFAA           HAVGGALFFAA 0 60 0 120 SSALGKVLWDG
40	301 <u>LIG</u> 351 <u>FAG</u> 401 QSL  ORF125ng-1 and Of  orf125-1.pep  orf125ng-1	TVLAVML LVLWLAG QRNPS* RF125-1 MSGNAS             MSGNAS	PVTEYKNFLL FILYRFLLSS  show 95.19  10 SSPSSSSAIGL:              SSPSSSAAIGLY 10  70 LTGRSSMESVR:	GWESSIGLTA  6 identity in  20 IWFGAAVSIAE:          VWFGAAVSIAE: 20  80 LSFGKRGSVLF:	PVMSAVAIAT  408 aa overl  30 ISTGTLLAPLGI ISTGTLLAPLGI 30  90 ISVANMLQLAGWI	VSVRLFFKKT lap: 40 5 WQRGLAALLLG IIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	0 60 HAVGGALFFAA IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
40	301 <u>LIG</u> 351 <u>FAG</u> 401 QSL  ORF125ng-1 and Of  orf125-1.pep  orf125ng-1	TVLAVML LVLWLAG QRNPS* RF125-1 MSGNAS             MSGNAS	PVTEYKNFLL FILYRFLLSS  show 95.19  10 SSPSSSSAIGL:              SSPSSSAAIGLY 10  70 LTGRSSMESVR:	GWESSIGLTA  6 identity in  20 IWFGAAVSIAE:          VWFGAAVSIAE: 20  80 LSFGKRGSVLF:	PVMSAVAIAT  408 aa overl  30 ISTGTLLAPLGI ISTGTLLAPLGI 30  90 ISVANMLQLAGWI	VSVRLFFKKT    ap:   40	0 60 HAVGGALFFAA IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
40 45	301 <u>LIG</u> 351 <u>FAG</u> 401 QSL  ORF125ng-1 and Of  orf125-1.pep  orf125ng-1	TVLAVML LVLWLAG QRNPS* RF125-1 MSGNAS             MSGNAS	PVTEYKNFLL FILYRFLESS  show 95.19  10 SSPSSSSAIGL            SSPSSSAAIGL 10  70 LTGRSSMESVR            LTGRSSMESVR	GWESSIGLTA  6 identity in  20 IWFGAAVSIAE:          WWFGAAVSIAE: 20  80 LSFGKRGSVLF:           LSFGKCGSVLF:	PVMSAVAIAT  408 aa overl  30 ISTGTLLAPLGI ISTGTLLAPLGI 30 90 10 SVANMLQLAGW 111111111111111111111111111111111111	VSVRLFFKKT lap: 40 5 WQRGLAALLG WQRGLAALLG 40 5 11 TAVMIYAGATV      :     TAVMIYVGATV 00 11	0 60 HAVGGALFFAA           HAVGGALFFAA 0 60 0 120 SSALGKVLWDG
40	301 <u>LIG</u> 351 <u>FAG</u> 401 QSL  ORF125ng-1 and O  orf125-1.pep orf125ng-1  orf125-1.pep orf125ng-1	TVLAVML LVLWLAG QRNPS*  RF125-1  MSGNAS       MSGNAS AYIGAI       AYIGAI	PVTEYKNFLL FILYRFLESS  show 95.19  10 SSPSSSSAIGL            SSPSSSAAIGL 10  70 LTGRSSMESVR:            LTGRSSMESVR: 70  130	GWESSIGLTA  6 identity in  20 IWFGAAVSIAE:          WWFGAAVSIAE: 20  80 LSFGKRGSVLF:            LSFGKCGSVLF: 80  140	PVMSAVAIAT  408 aa overl  30 ISTGTLLAPLGI ISTGTLLAPLGI 30 90 10 SVANMLQLAGW 111111111111111111111111111111111111	VSVRLFFKKT lap: 40 5 WQRGLAALLIG WQRGLAALLIG 40 5 00 11 TAVMIYAGATV            TAVMIYVGATV 00 11	0 60 HAVGGALFFAA            HAVGGALFFAA 0 60 0 120 SSALGKVLWDG            SSALGKVLWDG 0 120
40 45	301 <u>LIG</u> 351 <u>FAG</u> 401 QSL  ORF125ng-1 and Of  orf125-1.pep  orf125ng-1	TVLAVML LVLWLAG QRNPS*  RF125-1  MSGNAS IIIII MSGNAS  AYIGAI IIIII AYIGAI	PVTEYKNFLL FILYRFLLSS  show 95.19  10 SSPSSSSAIGLE                   SSPSSSAAIGLE 10  70 LTGRSSMESVRE 11 LTGRSSMESVRE 70  130 WALANGALIVLE	GWESSIGLTA  6 identity in  20 IWFGAAVSIAE: :          WWFGAAVSIAE: 20  80 LSFGKRGSVLF: 80 140 WLVFGARKTGG	PVMSAVAIAT  408 aa overl  30 ISTGTLLAPLGI 30 90 10 SVANMLQLAGW 11        SVANMLQLAGW 90 10 150 150 16 LKTVSMLLMLLI	VSVRLFFKKT    Ap:   40	0 60 HAVGGALFFAA IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
40 45	301 LIG 351 FAG 401 QSL ORF125ng-1 and Or orf125-1.pep orf125ng-1 orf125-1.pep orf125ng-1	TVLAVML LVLWLAG QRNPS* RF125-1  MSGNAS IIIII MSGNAS  AYIGAI IIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	PVTEYKNFLL FILYRFLLSS  show 95.19  10 SSPSSSSAIGLE                       SSPSSSAAIGLE 10  70 LTGRSSMESVRE                     LTGRSSMESVRE 70  130 WALANGALIVLE	GWESSIGLTA  6 identity in  20 IWFGAAVSIAE:            WWFGAAVSIAE: 20  80 LSFGKRGSVLF: 80  140 WLVFGARKTGG:	PVMSAVAIAT  408 aa overl  30 ISTGTLLAPLGI ISTGTLLAPLGI 30  90 10 SVANMLQLAGW 111111111111111111111111111111111111	VSVRLFFKKT    Ap:   40	0 60 HAVGGALFFAA IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
40 45	301 <u>LIG</u> 351 <u>FAG</u> 401 QSL  ORF125ng-1 and O  orf125-1.pep orf125ng-1  orf125-1.pep orf125ng-1	TVLAVML LVLWLAG QRNPS* RF125-1  MSGNAS IIIII MSGNAS  AYIGAI IIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	PVTEYKNFLL FILYRFLLSS  show 95.19  10 SSPSSSSAIGLE                       SSPSSSAAIGLE 10  70 LTGRSSMESVRE                     LTGRSSMESVRE 70  130 WALANGALIVLE	GWESSIGLTA  6 identity in  20 IWFGAAVSIAE:            WWFGAAVSIAE: 20  80 LSFGKRGSVLF: 80  140 WLVFGARKTGG:	PVMSAVAIAT  408 aa overl  30 ISTGTLLAPLGI ISTGTLLAPLGI 30  90 10 SVANMLQLAGW 111111111111111111111111111111111111	VSVRLFFKKT    Ap:   40	0 60 HAVGGALFFAA IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
40 45 50	301 LIG 351 FAG 401 QSL ORF125ng-1 and Or orf125-1.pep orf125ng-1 orf125-1.pep orf125ng-1	TVLAVML LVLWLAG QRNPS* RF125-1  MSGNAS IIIII MSGNAS  AYIGAI IIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	PVTEYKNFLL FILYRFLLSS  show 95.19  10 SSPSSSSAIGLE                       SSPSSSAAIGLE 10  70 LTGRSSMESVRE                     LTGRSSMESVRE 70  130 WALANGALIVLE	GWESSIGLTA  6 identity in  20 IWFGAAVSIAE:            WWFGAAVSIAE: 20  80 LSFGKRGSVLF: 80  140 WLVFGARKTGG:	PVMSAVAIAT  408 aa overl  30 ISTGTLLAPLGI ISTGTLLAPLGI 30  90 10 SVANMLQLAGW 11         SVANMLQLAGW 90 10 LKTVSMLLMLLL LKTVSMLLMLLL LKTVSMLLMLLL	VSVRLFFKKT    Ap:   40	0 60 HAVGGALFFAA IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
40 45 50	301 LIG 351 FAG 401 QSL ORF125ng-1 and Or orf125-1.pep orf125ng-1 orf125-1.pep orf125ng-1	TVLAVML LVLWLAG QRNPS* RF125-1  MSGNAS IIIII MSGNAS  AYIGAI IIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	PVTEYKNFLL FILYRFLLSS  show 95.19  10 SSPSSSSAIGLE                       SSPSSSAAIGLE 10  70 LTGRSSMESVRE 10  130 WALANGALIVLE                       WALANGALIVLE   WALANGALIVLE	GWESSIGLTA  6 identity in  20 IWFGAAVSIAE:            WWFGAAVSIAE: 20  80 LSFGKRGSVLF: 80  140 WLVFGARKTGG:	PVMSAVAIAT  408 aa overl  30 ISTGTLLAPLGI ISTGTLLAPLGI 30 90 10 SVANMLQLAGW 90 11 ISVANMLQLAGW 90 15 LKTVSMLLMLLL LKTVSMLLMLLL LKTVSMLLMLLL	VSVRLFFKKT    Ap:   40	0 60 HAVGGALFFAA IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
40 45	301 LIG 351 FAG 401 QSL ORF125ng-1 and Or orf125-1.pep orf125ng-1 orf125-1.pep orf125ng-1 orf125-1.pep orf125ng-1	TVLAVML LVLWLAG QRNPS* RF125-1  MSGNAS IIIII MSGNAS AYIGAI IIIII AYIGAI ESFVWI	PVTEYKNFLL FILYRFLLSS  show 95.19  10 SSPSSSSAIGL:                SSPSSSAAIGLY 10  70 LTGRSSMESVR: 70  130 WALANGALIVLY WALANGALIVLY 130	GWESSIGLTA  6 identity in  20 IWFGAAVSIAE:           WWFGAAVSIAE: 20  80 LSFGKRGSVLF: 80  140 WLVFGARKTGG:	PVMSAVAIAT  408 aa overl  30 ISTGTLLAPLGI ISTGTLLAPLGI 30  90 ISVANMLQLAGWI ISVANMLQLAGWI SVANMLQLAGWI SVANMLQLAGWI LKTVSMLLMLLL LKTVSMLLMLLL LKTVSMLLMLLL LKTVSMLLMLLL LKTVSMLLMLLL LKTVSMLLMLLL LKTVSMLLMLLL LSO 1	VSVRLFFKKT    Ap:   40	0 60 HAVGGALFFAA IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
40 45 50	301 LIG 351 FAG 401 QSL ORF125ng-1 and Off orf125-1.pep orf125ng-1 orf125-1.pep orf125ng-1 orf125-1.pep orf125ng-1	TVLAVML LVLWLAG QRNPS* RF125-1  MSGNAS       MSGNAS AYIGAI       AYIGAI ESFVWI	PVTEYKNFLL FILYRFLLSS  show 95.19  10 SSPSSSSAIGL:              SSPSSSAAIGL:               SSPSSSAAIGL: 10  70 LTGRSSMESVR: 70  130 WALANGALIVL!             WALANGALIVL! 130  190	GWESSIGLTA  6 identity in  20 IWFGAAVSIAE:           WWFGAAVSIAE: 20  80 LSFGKRGSVLF: 80  140 WLVFGARKTGG:            WLVFGARKTGG: 140 200	PVMSAVAIAT  408 aa overl  30 ISTGTLLAPLGI ISTGTLLAPLGI 30 90 SVANMLQLAGW 111111111111111111111111111111111111	VSVRLFFKKT lap: 40 5 WQRGLAALLLG IIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	0 60 HAVGGALFFAA IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
40 45 50	301 LIG 351 FAG 401 QSL ORF125ng-1 and Or orf125-1.pep orf125ng-1 orf125-1.pep orf125ng-1 orf125-1.pep orf125ng-1	TVLAVML LVLWLAG QRNPS* RF125-1  MSGNAS IIIII MSGNAS AYIGAI IIIII AYIGAI ESFVWI IIIII ESFVWI 180 DGMSF0	PVTEYKNFLL FILYRFLLSS  show 95.19  10 SSPSSSSAIGL:                SSPSSSSAIGL:                 SSPSSSAAIGL' 10  70 LTGRSSMESVR: 70  130 WALANGALIVL' 130  190 GTAVELSAVMP:	GWESSIGLTA  6 identity in  20 IWFGAAVSIAE:           WWFGAAVSIAE: 20  80 LSFGKRGSVLF: 80  140 WLVFGARKTGG:             WLVFGARKTGG: 140 200 LSWLPLAADYT:	PVMSAVAIAT  408 aa overl  30 ISTGTLLAPLGI IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	VSVRLFFKKT lap:  40 5 WQRGLAALLG             WQRGLAALLLG 40 5 00 11 FAVMIYAGATV            FAVMIYVGATV 00 11 60 17 AVLWLSAEVFS            AVLWLSVEVFA 60 17 220 2 TATLAYTLTGG	0 60 HAVGGALFFAA IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
40 45 50	301 LIG 351 FAC 401 QSL ORF125ng-1 and O orf125-1.pep orf125ng-1 orf125-1.pep orf125ng-1 orf125-1.pep orf125ng-1	TVLAVML LVLWLAG QRNPS* RF125-1  MSGNAS       MSGNAS AYIGAI       AYIGAI AYIGAI ESFVWI       ESFVWI	PVTEYKNFLL FILYRFLLSS  show 95.19  10 SSPSSSSAIGL:              SSPSSSAAIGL:              SSPSSSAAIGL: 10  70 LTGRSSMESVR: 70  130 WALANGALIVL!            WALANGALIVL! 130  190 STAVELSAVMP:	GWESSIGLTA  6 identity in  20 IWFGAAVSIAE:           IWFGAAVSIAE: 20  80 LSFGKRGSVLF: 80  140 WLVFGARKTGG:            WLVFGARKTGG: 140  200 LSWLPLAADYT:	PVMSAVAIAT  408 aa overl  30 ISTGTLLAPLGI ISTGTLLAPLGI 30 90 SVANMLQLAGW 90 11 ISTGTLLAPLGI ISTGTLLAPLGI 11 ISTGTLLAPLGI 10 ISTGTLAPLGI 10 ISTGTLAPLG	VSVRLFFKKT lap: 40 5 WQRGLAALLG IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	0 60 HAVGGALFFAA IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
40 45 50	301 LIG 351 FAG 401 QSL ORF125ng-1 and Off orf125-1.pep orf125ng-1 orf125-1.pep orf125ng-1 orf125-1.pep orf125ng-1	TVLAVML LVLWLAG QRNPS* RF125-1  MSGNAS       MSGNAS AYIGAI       AYIGAI AYIGAI ESFVWI       ESFVWI	PVTEYKNFLL FILYRFLLSS  show 95.19  10 SSPSSSSAIGL:              SSPSSSAAIGL:              SSPSSSAAIGL: 10  70 LTGRSSMESVR: 70  130 WALANGALIVL!                     130  190 GTAVELSAVMP:              GTAVELSAVMP:	GWESSIGLTA  6 identity in  20 IWFGAAVSIAE:           IWFGAAVSIAE: 20  80 LSFGKRGSVLF: 80  140 MLVFGARKTGG:            WLVFGARKTGG: 140  200 LSWLPLAADYT:	PVMSAVAIAT  408 aa overl  30 ISTGTLLAPLGI ISTGTLLAPLGI 30 90 SVANMLQLAGW 90 11 SVANMLQLAGW 90 15 LKTVSMLLMLL IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	VSVRLFFKKT lap: 40 5 WQRGLAALLG             WQRGLAALLG 40 5 00 11 TAVMIYAGATV              TAVMIYVGATV 00 11 60 17 AVLWLSAEVFS             AVLWLSVEVFA 60 17 220 2 TATLAYTLTGC	0 60 HAVGGALFFAA IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
40 45 50	301 LIG 351 FAC 401 QSL ORF125ng-1 and O orf125-1.pep orf125ng-1 orf125-1.pep orf125ng-1 orf125-1.pep orf125ng-1	TVLAVML LVLWLAG QRNPS* RF125-1  MSGNAS       MSGNAS AYIGAI       AYIGAI AYIGAI ESFVWI       ESFVWI	PVTEYKNFLL FILYRFLLSS  show 95.19  10 SSPSSSSAIGL:              SSPSSSAAIGL:              SSPSSSAAIGL: 10  70 LTGRSSMESVR: 70  130 WALANGALIVL!            WALANGALIVL! 130  190 STAVELSAVMP:	GWESSIGLTA  6 identity in  20 IWFGAAVSIAE:           IWFGAAVSIAE: 20  80 LSFGKRGSVLF: 80  140 MLVFGARKTGG:            WLVFGARKTGG: 140  200 LSWLPLAADYT:	PVMSAVAIAT  408 aa overl  30 ISTGTLLAPLGI ISTGTLLAPLGI 30 90 SVANMLQLAGW 90 11 SVANMLQLAGW 90 15 LKTVSMLLMLL IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	VSVRLFFKKT lap: 40 5 WQRGLAALLG IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	0 60 HAVGGALFFAA IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
40 45 50	301 LIG 351 FAC 401 QSL ORF125ng-1 and O orf125-1.pep orf125ng-1 orf125-1.pep orf125ng-1 orf125-1.pep orf125ng-1	TVLAVML LVLWLAG QRNPS* RF125-1  MSGNAS       MSGNAS AYIGAI       AYIGAI AYIGAI ESFVWI       ESFVWI	PVTEYKNFLL FILYRFLLSS  show 95.19  10 SSPSSSSAIGL:              SSPSSSAAIGL:              SSPSSSAAIGL: 10  70 LTGRSSMESVR: 70  130 WALANGALIVL!                     130  190 GTAVELSAVMP:              GTAVELSAVMP:	GWESSIGLTA  6 identity in  20 IWFGAAVSIAE:           IWFGAAVSIAE: 20  80 LSFGKRGSVLF: 80  140 MLVFGARKTGG:            WLVFGARKTGG: 140  200 LSWLPLAADYT:	PVMSAVAIAT  408 aa overl  30 ISTGTLLAPLGI ISTGTLLAPLGI 30 90 SVANMLQLAGW 90 11 SVANMLQLAGW 90 15 LKTVSMLLMLL IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	VSVRLFFKKT lap: 40 5 WQRGLAALLG             WQRGLAALLG 40 5 00 11 TAVMIYAGATV              TAVMIYVGATV 00 11 60 17 AVLWLSAEVFS             AVLWLSVEVFA 60 17 220 2 TATLAYTLTGC	0 60 HAVGGALFFAA IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
40 45 50	301 LIG 351 FAG 401 QSL ORF125ng-1 and Or orf125-1.pep orf125ng-1 orf125-1.pep orf125ng-1 orf125-1.pep orf125ng-1 orf125ng-1	TVLAVML LVLWLAG QRNPS* RF125-1  MSGNAS       MSGNAS AYIGAI       AYIGAI AYIGAI ESFVWI       ESFVWI	PVTEYKNFLL FILYRFLLSS  show 95.19  10 SSPSSSSAIGL:              SSPSSSAAIGL:              SSPSSSAAIGL: 10  70 LTGRSSMESVR: 70  130 WALANGALIVL!                     130  190 GTAVELSAVMP:              GTAVELSAVMP:	GWESSIGLTA  6 identity in  20 IWFGAAVSIAE:           IWFGAAVSIAE: 20  80 LSFGKRGSVLF: 80  140 MLVFGARKTGG:            WLVFGARKTGG: 140  200 LSWLPLAADYT:	PVMSAVAIAT  408 aa overl  30 ISTGTLLAPLGI ISTGTLLAPLGI 30 90 10 SVANMLQLAGWI 11         SVANMLQLAGWI 150 10 LKTVSMLLMLLI 150 11 LKTVSMLLMLLI 150 10 RHARRPFAATLI 11         1210 RHARRPFAATLI 11         1210 RHARRPFAATLI 1210 2210 2210 2210 2210 2210 2210 221	VSVRLFFKKT    Ap:   40	0 60 HAVGGALFFAA IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
40 45 50	301 LIG 351 FAG 401 QSL ORF125ng-1 and Or orf125-1.pep orf125ng-1 orf125-1.pep orf125ng-1 orf125-1.pep orf125ng-1 orf125ng-1	TVLAVML LVLWLAG QRNPS* RF125-1  MSGNAS IIIII MSGNAS AYIGAI IIIII AYIGAI ESFVWI IIIII ESFVWI 180 DGMSFC IIIIII DGMTFC	PVTEYKNFLL FILYRFLLSS  show 95.19  10 SSPSSSSAIGLE             SSPSSSAAIGLE 10  70 LTGRSSMESVRE 70  130 WALANGALIVLE            WALANGALIVLE	GWESSIGLTA  6 identity in  20 IWFGAAVSIAE: :           WFGAAVSIAE: 20  80 LSFGKRGSVLF: 80  140 WLVFGARKTGG:             WLVFGARRTGG: 140  200 LSWLPLAADYT: LSWLPLAADYT: LSWLPLAADYT: 200  260	PVMSAVAIAT  408 aa overl  30 ISTGTLLAPLGI 30 90 10 SVANMLQLAGW 11         SVANMLQLAGW 90 10 LKTVSMLLMLL 11        LKTVSMLLMLL 150 10 210 RHARRPFAATL 11         RQARRPFAATL 210 270	VSVRLFFKKT    ap:   40	0 60 HAVGGALFFAA              HAVGGALFFAA 0 60 0 120 SSALGKVLWDG              SSALGKVLWDG 0 120 0 179 TAGSTAAQ-VS :: ::      SSGTNAAPAVS 0 180                  WMYALGLAAAL                 WMYALGLAAAL                 WMYALGLAAAL
40 45 50	301 LIG 351 FAG 401 QSL ORF125ng-1 and Or orf125-1.pep orf125ng-1 orf125-1.pep orf125ng-1 orf125-1.pep orf125ng-1 orf125ng-1	TVLAVML LVLWLAG QRNPS* RF125-1  MSGNAS IIIII MSGNAS AYIGAI IIIII AYIGAI ESFVWI IIIII ESFVWI 180 DGMSF( IIIII DGMTF( 240 FTGET(	PVTEYKNFLL FILYRFLLSS  show 95.19  10 SSPSSSAIGLE              SSPSSSAAIGLE 10  70 LTGRSSMESVRE 10  130 WALANGALIVLE            WALANGALIVLE 130  190 STAVELSAVMPE 190  250 DVAKILLGAGLE	GWESSIGLTA  6 identity in  20 IWFGAAVSIAE:              WFGAAVSIAE: 20  80 LSFGKRGSVLF: 80  140 WLVFGARKTGG:               WLVFGARRTGG: 140  200 LSWLPLAADYT:            LSWLPLAADYT: 200  260 GAAGILAVVLS	PVMSAVAIAT  408 aa overl  30 ISTGTLLAPLGI 30 90 ISTGTLLAPLGI 30 90 ISVANMLQLAGWI ISVANMLQLAGWI ISVANMLQLAGWI ISVANMLQLAGWI 150 150 150 16 LKTVSMLLMLLI ISTGTLLAPLGI ISTGTLLAPLGI ISTGTLLAPLGI ISTGTLLAPLGI ISTGTLLAPLGI ISTGTLAPLGI ISTGTL	VSVRLFFKKT    Ap:   40	0 60 HAVGGALFFAA            HAVGGALFFAA 0 60 0 120 SSALGKVLWDG               SSALGKVLWDG 0 120 0 179 TAGSTAAQ-VS :: ::      SSGTNAAPAVS 0 180 30 239 WMYALGLAAAL             WMYALGLAAAL             WMYALGLAAAL 0 240
40 45 50 55	301 LIG 351 FAG 401 QSL ORF125ng-1 and Or orf125-1.pep orf125ng-1 orf125-1.pep orf125ng-1 orf125-1.pep orf125ng-1 orf125-1.pep orf125ng-1	TVLAVML LVLWLAG QRNPS* RF125-1  MSGNAS IIIII MSGNAS AYIGAI IIIII AYIGAI ESFVWI IIIII ESFVWI 180 DGMSFT IIIII DGMTFC 240 FTGETI	PVTEYKNFLL FILYRFLLSS  show 95.19  10 SSPSSSAAIGLE              SSPSSSAAIGLE 10  70 LTGRSSMESVRE 70  130 WALANGALIVLE 130 190 GTAVELSAVMP 191 STAVELSAVMP 190 250 DVAKILLGAGLE	GWESSIGLTA  6 identity in  20 IWFGAAVSIAE:             WFGAAVSIAE: 20  80 LSFGKRGSVLF: 80  140 WLVFGARKTGG:               WLVFGARRTGG: 140  200 LSWLPLAADYT:            LSWLPLAADYT: 200  260 GAAGILAVVLS:  :	PVMSAVAIAT  408 aa overl  30 ISTGTLLAPLGI 30 ISTGTLLAPLGI 30 90 10 SVANMLQLAGW	VSVRLFFKKT    Ap:	0 60 HAVGGALFFAA            HAVGGALFFAA 0 60 0 120 SSALGKVLWDG              SSALGKVLWDG 0 120 0 179 TAGSTAAQ-VS :: ::      SSGTNAAPAVS 0 180 30 239 WMYALGLAAAL             WMYALGLAAAL             WMYALGLAAAL 0 240
40 45 50	301 LIG 351 FAG 401 QSL ORF125ng-1 and Or orf125-1.pep orf125ng-1 orf125-1.pep orf125ng-1 orf125-1.pep orf125ng-1 orf125ng-1	TVLAVML LVLWLAG QRNPS* RF125-1  MSGNAS IIIII MSGNAS AYIGAI IIIII AYIGAI ESFVWI IIIII ESFVWI 180 DGMSFT IIIII DGMTFC 240 FTGETI	PVTEYKNFLL FILYRFLLSS  show 95.19  10 SSPSSSAAIGLE                       SSPSSSAAIGLE 10  70 LTGRSSMESVRE 70  130 WALANGALIVLE 130  190 GTAVELSAVMP 191 GTAVELSAVMP 190  250 DVAKILLGAGLE	GWESSIGLTA  6 identity in  20 IWFGAAVSIAE:              WFGAAVSIAE: 20  80 LSFGKRGSVLF: 80  140 WLVFGARKTGG:               WLVFGARRTGG: 140 200 LSWLPLAADYT:            LSWLPLAADYT: 200 260 GAAGILAVVLS:  :          GITGILAVVLS:	PVMSAVAIAT  408 aa overl  30 ISTGTLLAPLGI 30 ISTGTLLAPLGI 30 90 10 SVANMLQLAGW	VSVRLFFKKT    Ap:	0 60 HAVGGALFFAA            HAVGGALFFAA 0 60 0 120 SSALGKVLWDG              SSALGKVLWDG 0 120 0 179 TAGSTAAQ-VS :: ::      SSGTNAAPAVS 0 180 30 239 WMYALGLAAAL             WMYALGLAAAL             WMYALGLAAAL 0 240 190 299 EFAETPVAVGVT

```
300
                           310
                                   320
                                           330
                                                   340
                                                            350
                     LIGTVLAVMLPVTEYENFLLLIGSVFAPMAAVLIADFFVLKRREEIEGFDFAGLVLWLAG
        orf125-1.pep
                     5
        orf125ng-1
                     LIGTVLAVMLPVTEYKNFLLLIGSVFAPMAAVLIADFFVLKRREEIEGFDFAGLVLWLAG
                          310
                                  320
                                          330
                                                   340 .
                           370
                                   380
                                           390
                                                   400
                     FILYRFLLSSGWESSIGLTAPVMSAVAIATVSVRLFFKKTQSLQRNPSX
        orf125-1.pep
10
                     **||**|||
                     FILYRFLLSSGWESSIGLTAPVMSAVAIATVSVRLFFKKTQSLQRNPSX
        orf125ng-1
                           370
                                   380
                                           390
```

Based on this analysis, including the presence of putative leader sequence and transmembrane domains in the gonococcal protein, it is predicted that the proteins from N. meningitidis and 15 N.gonorrhoeae, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 96

751

801

851

901

951

1001

1051 1101

55

The following partial DNA sequence was identified in N. meningitidis <SEO ID 809>:

```
ATGACCCGTA TCGCCATCCT CGGCGGCGCC CTCTCGGGAA GGCTGACCGC
20
                     GTTGCAGCTT GCAGAACAAG GTTATCAGAT TGCACTTTTC GATAAAAGCT
                 51
                101
                     GCCGCCGGGG CGAACACGCC GCCGCCTATG TAGCCGCCGC CATGCTCGCG
                     CCTGCAGCGG A.ACGGTCGA AGCCACGCCC GAAGTGGTCA GGCTGGGCAG
                151
                201
                     GCAGAGCATC CCGCTTTGGC GCGGCATCCG ATGCCGTCTG AACACGCACA
                     CGATGATGCA GGAAAACGGC AGCCTGATTG TATGGCACGG GCAGGACAAG
                251
25
                301
                     CCATTATCCA GCGAGTTCGT CCGCCATCTC AAACGCGGCG GCGT.ACGGA
                     TGACGAAATC GTCCGTTGGC GCGCCGACGA CATCGCCGAA CGCGAACCGC
                351
                401
                     AACTCGGCGG ACGTTTTTAA GACGGCATCT ACCTGCCGAC CGAAGC.CAG
                     CTCGACGGC GGCAATTATA GTCTGCACTT GCCGACGCTT TGGACGAACT
                451
                501
                     GAACGTCCCC TGCCATTGGG AACACGAATG CGTCCCCGAA GCCTGCAAG..
     This corresponds to the amino acid sequence <SEQ ID 810; ORF126>:
30
                     MTRIAILGGG LSGRLTALQL AEQGYQIALF DKSCRRGEHA AAYVAAAMLA
                 51
                     PAAXTVEATP EVVRLGRQSI PLWRGIRCRL NTHTMMOENG SLIVWHGODK
                     PLSSEFVRHL KRGGXTDDEI VRWRADDIAE REPQLGGRFX DGIYLPTEXQ
                     LDGRQLXSAL ADALDELNVP CHWEHECVPE ACK...
35
     Further work revealed the complete nucleotide sequence <SEQ ID 811>:
                     ATGACCCGTA TCGCCATCCT CGGCGGCGGC CTCTCGGGAA GGCTGACCGC
                     GTTGCAGCTT GCAGAACAAG GTTATCAGAT TGCACTTTTC GATAAAGGCT
                 51
                101
                     GCCGCCGGGG CGAACACGCC GCCGCCTATG TTGCCGCCGC CATGCTCGCG
                     CCTGCGGCGG AAGCGGTCGA AGCCACGCCC GAAGTGGTCA GGCTGGGCAG
                151
40
                     GCAGAGCATC CCGCTTTGGC GCGGCATCCG ATGCCGTCTG AACACGCACA
                201
                     CGATGATGCA GGAAAACGGC AGCCTGATTG TGTGGCACGG GCAGGACAAG
                251
                     CCATTATCCA GCGAGTTCGT CCGCCATCTC AAACGCGGCG GCGTAGCGGA
                301
                     TGACGAAATC GTCCGTTGGC GCGCCGACGA CATCGCCGAA CGCGAACCGC
                351
                401
                     AACTCGGCGG ACGTTTTTCA GACGGCATCT ACCTGCCGAC CGAAGGCCAG
45
                     CTCGACGGC GGCAAATATT GTCTGCACTT GCCGACGCTT TGGACGAACT
                451
                501
                     GAACGTCCCC TGCCATTGGG AACACGAATG CGTCCCCGAA GGCCTGCAAG
                     CCCAATACGA CTGGCTGATC GACTGCCGCG GCTACGGCGC AAAAACCGCG
                551
                601
                     TGGAACCAAT CCCCCGAGCA CACCAGCACC CTGCGCGGCA TACGCGGCGA
                     AGTGGCGCGG GTTTACACAC CCGAAATCAC GCTCAACCGC CCCGTGCGTC
                651
50
                     TGCTCCATCC GCGTTATCCG CTCTACATCG CCCCGAAAGA AAACCACGTC
                701
```

TTCGTCATCG GCGCGACCCA AATCGAAAGC GAAAGCCAAG CCCCCGCCAG

CGTGCGTTCA GGGTTGGAAC TCTTGTCCGC ACTCTATGCC ATCCACCCCG CCTTCGGCGA AGCCGACATC CTCGAAATCG CCACCGGCCT GCGCCCCACG

CTCAACCACC ACAACCCCGA AATCCGTTAC AACCGCGCCC GACGCCTGAT

TGAAATCAAC GGCCTTTTCC GCCACGGTTT CATGATCTCC CCCGCCGTAA CCGCCGCCGC CGCCAGATTG GCAGTGGCAC TGTTTGACGG AAAAGACGCG

CCCGAACGCG ATAAAGAAAG CGGTTTGGCG TATATCCGAA GACAAGATTA

This corresponds to the amino acid sequence <SEQ ID 812; ORF126-1>:

```
MTRIAILGGG LSGRLTALQL AEQGYQIALF DKGCRRGEHA AAYVAAAMLA
                51
                    PAAEAVEATP EVVRLGRQSI PLWRGIRCRL NTHTMMQENG SLIVWHGODK
               101
                    PLSSEFVRHL KRGGVADDEI VRWRADDIAE REPOLGGRFS DGIYLPTEGO
5
               151
                    LDGRQILSAL ADALDELNVP CHWEHECVPE GLQAQYDWLI DCRGYGAKTA
                    WNQSPEHTST LRGIRGEVAR VYTPEITLNR PVRLLHPRYP LYIAPKENHV
               201
               251
                    FVIGATQIES ESQAPASVRS GLELLSALYA IHPAFGEADI LEIATGLRPT
                    LNHHNPEIRY NRARRLIEIN GLFRHGFMIS PAVTAAAARL AVALFDGKDA
               301
                    PERDKESGLA YIRRQD*
               351
```

10 Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from N. meningitidis (strain A)

ORF126 shows 90.0% identity over a 180aa overlap with an ORF (ORF126a) from strain A of N. meningitidis:

```
20
                                        30
                                                40
15
        orf126.pep
                  MTRIAILGGGLSGRLTALQLAEQGYQIALFDKSCRRGEHAAAYVAAAMLAPAAXTVEATP
                  orf126a
                  MTRIAILGGGLSGRLTALQLAEQGYQIALFDKGCRRGEHAAAYVAAAMLAPAAEAVEATP
                                20
                                        30
                                                40
                                                        50
20
                         70
                                80
                                        90
                                               100
                                                       110
                                                              120
        orf126.pep
                  EVVRLGRQSIPLWRGIRCRLNTHTMMQENGSLIVWHGQDKPLSSEFVRHLKRGGXTDDEI
                  EVVRLGRQXIPLWRGIRCHLKTPAMMXENGSLIVWHGQDKPLSNEFVRHLKRGGVADDXI
        orf126a
                         70
                                80
                                        90
                                               100
                                                       110
                                                              120
25
                        130
                                140
                                       150
                                               160
                                                       170
        orf126.pep
                  VRWRADDIAEREPQLGGRFXDGIYLPTEXQLDGRQLXSALADALDELNVPCHWEHECVPE
                   orf126a
                  VRWRADDIAEREPQLGGRFSDGIYLPTEGQLDGRQILSALADALDELNVPCHWEHECAPE
30
                        130
                                140
                                       150
                                               160
```

The complete length ORF126a nucleotide sequence <SEQ ID 813> is:

	1	ATGACCCGTA	TCGCCATCCT	CGGCGGCGGC	CTCTCNGGAA	GGCTGACCGC
	51	ACTGCAGCTT	GCAGAACAAG	GTTATCAGAT	TGCACTTTTC	GATAAAGGCT
	101	GCCGCCGGGG	CGAACACGCC	GCCGCCTATG	TTGCCGCCGC	CATGCTCGCG
35	151	CCTGCGGCGG	AAGCGGTCGA	AGCCACGCCT	GAAGTGGTCA	GGCTGGGCAG
	201	GCAGANCATC	CCGCTTTGGC	GCGGCATCCG	ATGCCATCTG	AAAACGCCTG
	251	CCATGATGCA	NGAAAACGGC	AGCCTGATTG	TGTGGCACGG	GCAGGACAAA
	301	CCTTTATCCA	ACGAGTTCGT	CCGCCATCTC	AAACGCGGCG	GCGTAGCGGA
	351	TGACNAAATC	GTCCGTTGGC	GCGCCGACGA	CATCGCCGAA	CGCGAACCGC
40	401	AACTCGGCGG	ACGTTTTTCA	GACGGCATCT	ACCTGCCGAC	CGAAGGCCAG
	451	CTCGACGGGC	GGCAAATATT	GTCTGCACTT	GCCGACGCTT	TGGACGAACT
	501	GAACGTCCCC	TGCCATTGGG	AACACGAATG	TGCCCCCGAA	GACTTGCAAG
	551	CCCAATACGA	CTGGCTGATC	GACTGCCGCG	GCTACGGCGC	AAAAACCGCG
	601	TGGAACCAAT	CCCCGANNA	NACCAGCACC	CTGCGCGGCA	TACGCGGCGA
45	651	AGTGGCGCGG	GTTTACACAC	CCGAAATCAC	GCTCAACCGC	CCCGTGCGCC
	701	TGCTACACCC	GCGCTATCCG	CTNTACATCG	CCCCGAAAGA	AAACCNCGTC
	751	TTCGTCATCG	GCGCGACCCA	AATCGAAAGC	GAAAGCCAAG	CACCTGCCAG
	801	CGTGCGTTCC	GGGCTGGAAC	TCTTATCCGC	ACTCTATGCC	GTCCACCCCG
	851	CCTTCGGCGA	AGCCGACATC	CTCGAAATCG	CCACCGGCCT	GCGCCCCACG
50	901	CTCAATCACC	ACAACCCCGA	AATCCGTTAC	AACCGCGCCC	GACGCCTGAT
	951	TGAAATCAAC	GGCCTTTTCC	GCCACGGTTT	CATGATCTCC	CCCGCCGTAA
	1001	CCGCCGCCGC	CGTCAGATTG	GCAGTGGCAC	TGTTTGACGG	AAAAGANGCG
	1051	CCCGAACGCG	ATGAAGAAAG	CGGTTTGGCG	TATATCCGAA	GACAAGATTA
	1101	A				

55 This encodes a protein having amino acid sequence <SEO ID 814>:

60

1	MTRIAILGGG	LSGRLTALQL	<b>AEQGYQIALF</b>	DKGCRRGEHA	AAYVAAAMLA
51	PAAEAVEATP	EVVRLGRQXI	PLWRGIRCHL	KTPAMMXENG	SLIVWHGQDK
101	PLSNEFVRHL	KRGGVADDXI	VRWRADDIAE	REPOLGGRES	DGIYLPTEGQ
151	LDGRQILSAL	ADALDELNVP	CHWEHECAPE	DLQAQYDWLI	DCRGYGAKTA
201	WNQSPXXTST	LRGIRGEVAR	VYTPEITLNR	PVRLLHPRYP	LYIAPKENXV

251 FVIGATQIES ESQAPASVRS GLELLSALYA VHPAFGEADI LEIATGLRPT 301 LNHHNPEIRY NRARRLIEIN GLFRHGFM<u>IS PAVTAAAVRL AVALF</u>DGKXA 351 PERDEESGLA YIRRQD\*

# ORF126a and ORF126-1 show 95.4% identity in 366 aa overlap:

5	•	10	20	30	40	50	60
	orf126a.pep	MTRIAILGGGLSGRLTAI	QLAEQGYQ1	CALFDKGCRR	SEHAAAYVAA <i>I</i>	AMLAPAAEAVI	EATP
		111111111111111111111111111111111111111					1111
	orf126-1	MTRIAILGGGLSGRLTAI 10	QLAEQGYQI 20				
10		1,0	20	30	40	50	60
10		70	80	90	100	110	120
	orf126a.pep	EVVRLGRQXIPLWRGIRG					
	• •		:::::::				111
	orf126-1	EVVRLGRQSIPLWRGIRO	CRLNTHTMM	DENGSLIVWHO	ODKPLSSEF	/RHLKRGGVAI	DDEI
15		70	80	90	100	110	120
		100					
	orf126a.pep		L40	150	160	170	180
	Ollizog.pep	VRWRADDIAEREPQLGGF	(LODGIIPE)	LEGÖTDGKÖTI	SALADALDEI	LNVPCHWEHE	CAPE
20	orf126-1	VRWRADDIAEREPQLGGR	RESDCTYT.P1	I I I I I I I I I I I I I I I I I I I			
	<b>VIII</b> -	130 1	40	150	160	170	180
	•						
			200	210	220	230	240
25	orf126a.pep	DLQAQYDWLIDCRGYGAR	(TAWNQSPX)				
23	orf126-1	CI ONOVINI I DODOVONI		11111111		11111111	
	011126-1	GLQAQYDWLIDCRGYGAR 190 2	CTAWNQSPER	ATSTLEGIEGE 210	EVARVYTPEI: 220	LNRPVRLLHI 230	PRYP 240
		150 2	.00	210	220	230	240
		250 2	260	270	280	290	300
30	orf126a.pep	LYIAPKENXVFVIGATQI	(ESESQAPAS	SVRSGLELLS	ALYAVHPAFGI	EADILEIATG	LRPT
			[3][[]]	шинц			Ш
	orf126-1	LYIAPKENHVFVIGATQI	IESESQAPAS				
		250 2	260	270	280	290	300
35		310 3	320	330	340	350	360
	orf126a.pep	LNHHNPEIRYNRARRLIE				SKXAPERDEE!	SGLA
				ППППП	1:1111111	1: 11111:1	1111
	orf126-1	LNHHNPEIRYNRARRLIE					
40		310 3	320	330	340	350	360
40							
	orf126a.pep	YIRRODX					
	J22244. POP	111111					
	orf126-1	YIRRODX					
45							

#### Homology with a predicted ORF from N.gonorrhoeae

ORF126 shows 90% identity over a 180 aa overlap with a predicted ORF (ORF126ng) from N.gonorrhoeae:

50	orf126.pep	MTRIAILGGGLSGRLTALQLAEQGYQIALFDKSCRRGEHAAAYVAAAMLAPAAXTVEATP	60
	orf126ng	MTRIAVLGGGLSGRLTALQLAEQGYQIELFDKGTRQGEHAAAYVAAAMLAPAAEAVEATP	60
	orf126.pep	EVVRLGRQSIPLWRGIRCRLNTHTMMQENGSLIVWHGQDKPLSSEFVRHLKRGGXTDDEI	120
55	orf126ng	EVIRLGRQSIPLWRGIRCRLNTLTMMQENGSLIVWHGQDKPLSSEFVRHLKRGGVADDEI	120
	orf126.pep	VRWRADDIAEREPQLGGRFXDGIYLPTEXQLDGRQLXSALADALDEINVPCHWEHECVPE	180
	orf126ng	VRWRADEIAEREPQLGGRFSDGIYLPTEGQLDGRQILSALADALDELNVPCHWEHECAPQ	180

An ORF126ng nucleotide sequence <SEQ ID 815> was predicted to encode a protein having amino 60 acid sequence <SEQ ID 816>:

<sup>1</sup> MTRIAVLGGG LSGRLTALQL AEQGYQIELF DKGTRQGEHA AAYVAAAMLA

PCT/IB98/01665

```
PAAEAVEATP EVIRLGRQSI PLWRGIRCRL NTLTMMQENG SLIVWHGQDK
                     PLSSEFVRHL KRGGVADDEI VRWRADEIAE REPQLGGRFS DGIYLPTEGQ
                101
                     LDGRQILSAL ADALDELNVP CHWEHECAPQ DLQAQYDWVI DCRGYGAKTA
                151
                     WNQSPEHTST LRGIRGEVRG FTRPKSRSTA PCACCTRAIR STSPRKKTTS
                201
                     SSSARPKSKA KAKPPPAYVP GWNSYPRSMP STPPSAKPTS SKWRPGLRPT
5
                251
                     LNHHNPEIRY SRERRLIEIN GLFRHGFMIS PAVTAAAVRL AVALFDGKDA
                301
                     PERDEESGLA YIGRQD*
                351
     Further work revealed the following gonococcal DNA sequence <SEQ ID 817>:
                     ATGACCCGTA TCGCCGTCCT CGGAGGCGGC CTTTCCGGAA GGCTGACCGC
                     ATTGCAGCTT GCAGAACAAG GTTATCAGAT TGAACTTTTC GACAAGGGCA
10
                 51
                      CCCGCCAAGG CGAACACGCC GCCGCCTATG TTGCCGCCGC GATGCTCGCG
                101
                      CCTGCGGCGG AAGCGGTCGA GGCAACGCCC GAAGTCATCA GGCTGGGCAG
                151
                      GCAGAGCATT CCGCTTTGGC GCGGCATCCG ATGCCGTCTG AACACGCTCA
CGATGATGCA GGAAAACGGC AGCCTGATTG TGTGGCACGG GCAGGACAAG
                201
                251
                      CCATTATCCA GCGAGTTCGT CCGCCATCTC AAACGCGGCG GCGTAGCGGA
15
                301
                      TGACGAAATC GTCCGTTGGC GCGCCGATGA AATCGCCGAA CGCGAACCGC
                351
                      AACTCGGCGG ACGTTTTTCA GACGGCATCT ACCTGCCGAC CGAAGGCCAG
                 401
                      CTCGACGGC GGCAAATATT GTCTGCACTT GCCGACGCTT TGGACGAACT
                 451
                      GAACGTCCCT TGCCATTGGG AACACGAATG CGCCCCCCAA GACCTGCAAG
                 501
                      CCCAATACGA CTGGGTAATC GACTGCCGGG GCTACGGCGC GAAAACCGCG
TGGAACCAAT CCCCCGAGCA CACCAGCACC TTGCGCGGCA TACGCGGCGA
20
                 551
                 601
                      AGTGGCGCGG GTTTACACGC CCGAAATCAC GCTCAACCGC CCCGTGCGCC
                 651
                      TGCTGCACCC GCGCTATCCG CTCTACATCG CCCCGAAAGA AAACCACGTC
                 701
                      TTCGTCATCG GCGCGACCCA AATCGAAAGC GAAAGCCAAG CCCCCGCCAG
                 751
                      CGTACGTTCC GGGCTGGAAC TCTTATCCGC GCTCTATGCC GTCCACCCCG
CCTTCGGCGA AGCCGACATC CTCGAAATCG CCGCCGGCCT GCGCCCCACG
25
                 801
                 851
                      CTCAACCACC ACAACCCCGA AATCCGCTAC AGCCGCGAAC GCCGCCTCAT
                 901
                      CGAAATCAAC GGCCTTTTCC GGCACGGCTT TATGATTTCC CCCGCCGTAA
                 951
                      CCGCCGCCGC CGTCAGATTG GCAGTGGCAC TGTTTGACGG AAAAGACGCG
                1001
                      CCCGAACGTG ATGAAGAAAG CGGTTTGGCG TATATCGGAA GACAAGATTA
 30
                1051
                1101 A
      This corresponds to the amino acid sequence <SEQ ID 818; ORF126ng-1>:
                      MTRIAVLGGG LSGRLTALQL AEQGYQIELF DKGTRQGEHA AAYVAAAMLA
                      PAAEAVEATP EVIRLGRQSI PLWRGIRCRL NTLTMMQENG SLIVWHGQDK
                  51
                      PLSSEFVRHL KRGGVADDEI VRWRADEIAE REPQLGGRFS DGIYLPTEGQ
 35
                      LDGRQILSAL ADALDELNVP CHWEHECAPQ DLQAQYDWVI DCRGYGAKTA
                 151
                      WNQSPEHTST LRGIRGEVAR VYTPEITLNR PVRLLHPRYP LYIAPKENHV
                 201
                      FVIGATQIES ESQAPASVRS GLELLSALYA VHPAFGEADI LEIAAGLRPT
                 251
                      LNHHNPEIRY SRERRLIEIN GLFRHGFMIS PAVTAAAVRL AVALFDGKDA
                 301
                 351 PERDEESGLA YIGRQD*
 40
       ORF126ng-1 and ORF126-1 show 95.1% identity in 366 aa overlap:
                                                                   40
                                                                             50
                            MTRIAILGGGLSGRLTALQLAEQGYQIALFDKGCRRGEHAAAYVAAAMLAPAAEAVEATP
            orf126-1.pep
                            MTRIAVLGGGLSGRLTALQLAEQGYQIELFDKGTRQGEHAAAYVAAAMLAPAAEAVEATP
 45
            orf126ng-1
                                    10
                                              20
                                                         30
                                                                   40
                                                                             50
                                                                                        60
                                                         90
                                                                  100
                                                                            110
                            EVVRLGRQSIPLWRGIRCRLNTHTMMQENGSLIVWHGQDKPLSSEFVRHLKRGGVADDEI
             orf126-1.pep
                            H: HILLIAN HIL
 50
                            EVIRLGRQSIPLWRGIRCRLNTLTMMQENGSLIVWHGQDKPLSSEFVRHLKRGGVADDEI
             orf126ng-1
                                                         90
                                                                  100
                                                                             110
                                    70
                                               80
                                                                  160
                                   130
                                              140
                            VRWRADDIAEREPOLGGRFSDGIYLPTEGOLDGRQILSALADALDELNVPCHWEHECVPE
  55
             orf126-1.pep
                            VRWRADEIAEREPQLGGRFSDGIYLPTEGQLDGRQILSALADALDELNVPCHWEHECAPQ
             orf126ng-1
                                   130
                                                        150
                                                                  160
                                                                             170
                                                                                       180
                                              140
                                                        210
                                                                  220
                                                                             230
                                              200
  60
                            GLQAQYDWLIDCRGYGAKTAWNQSPEHTSTLRGIRGEVARVYTPEITLNRPVRLLHPRYP
             orf126-1.pep
                              DLQAQYDWVIDCRGYGAKTAWNQSPEHTSTLRGIRGEVARVYTPEITLNRPVRLLHPRYP
             orf126ng-1
                                                                  220
                                                                             230
                                    190
                                              200
                                                        210
  65
```

```
250
                                       260
                                                270
                                                         280
                       {\tt LYIAPKENHVFVIGATQIESESQAPASVRSGLELLSALYAIHPAFGEADILEIATGLRPT}
         orf126-1.pep
                       LYIAPKENHVFVIGATQIESESQAPASVRSGLELLSALYAVHPAFGEADILEIAAGLRPT
         orf126ng-1
5
                                       260
                                                270
                                                         280
                                                                  290
                              310
                                       320
                                                330
                                                         340
                                                                  350
                                                                           360
                       LNHHNPEIRYNRARRLIEINGLFRHGFMISPAVTAAAARLAVALFDGKDAPERDKESGLA
         orf126-1.pep
                       10
         orf126ng-1
                       LNHHNPEIRYSRERRLIEINGLFRHGFMISPAVTAAAVRLAVALFDGKDAPERDEESGLA
                                       320
                                                330
                                                         340
                                                                  350
         orf126-1.pep
                       YIRRQDX
15
                       11 1111
         orf126ng-1
                       YIGRQDX
     Furthermore, ORF126ng-1 shows homology to a putative Rhizobium oxidase flavoprotein:
          gi|2627327 (AF004408) putative amino acid oxidase flavoprotein [Rhizobium etli]
          Length = 327
20
          Score = 169 bits (423), Expect = 3e-41
          Identities = 112/329 (34%), Positives = 163/329 (49%), Gaps = 25/329 (7%)
                    Query: 3
                    RI V G G++G
                                A QL
                                      G+++ L ++
25
                    RILVNGAGVAGLTVAWQLYRHGFRVTLAERAGTVGA-GASGFAGGMLAPWCERESAEEPV 60
          Sbjct: 2
                    IRLGRQSIPLWRGIRCRLNTLTMMQENGSLIVWHGQDKPLSSEFVRHLKRGGVADDEIVR 122
          Query: 63
                    + LGR +
                                         +
                                            G+L+V G+D
                                                           F R
                                                                 G
                   LTLGRLAADWWEAA----LPGHVHRRGTLVVAGGRDTGELDRFSRRTS-GWEWLDEVA- 113
30
          Query: 123 WRADEIAEREPQLGGRFSDGIYLPTEGQLDGRQILSALADALDELNVPCHWEHECAPQDL 182
                        IA EP L GRF ++ E LD RO L+ALA L++ +
          Sbjct: 114 ----IAALEPDLAGRFRRALFFRQEAHLDPRQALAALAAGLEDARMRLTLG---VVGES 165
35
          Query: 183 QAQYDWVIDCRGYGAKTAWNQSPEHTSTLRGIRGEVARVYTPEITLNRPVRLLHPRYPLY 242
                      +D V+DC G
                                             LRG+RGE+ V T E++L+RPVRLLHPR+P+Y
          Sbjct: 166 DVDHDRVVDCTGAA-
                                       QIGRLPGLRGVRGEMLCVETTEVSLSRPVRLLHPRHPIY 218
          Query: 243 IAPKENHVFVIGATQIESESQAPASVRSGLELLSALYAVHPAFGEADILEIAAGLRPTLN 302
40
                    I P++ + F++GAT IES+
                                       P + RS +ELL+A YA+HPAFGEA + E AG+RP
          Sbjct: 219 IVPRDKNRFMVGATMIESDDGGPITARSLMELLNAAYAMHPAFGEARVTETGAGVRPAYP 278
          Query: 303 HHNPEIRYSRERRLIEINGLFRHGFMISP 331
                     + P R ++E R + +NGL+RHGF+++P
45
          Sbjct: 279 DNLP--RVTQEGRTLHVNGLYRHGFLLAP 305
```

This analysis suggests that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 97

The following DNA sequence, believed to be complete, was identified in N.meningitidis <SEQ ID

50 819>:

```
ATGACTGATA ATCGGGGGTT TACGCTGGTT GAATTAATAT CAGTGGTCTT
                51
                    GATATTGTCT GTACTTGCTT TAATTGTTTA TCCGAGCTAT CGCAATTATG
               101
                    TTGAGAAAGC AAAGATAAAT GCAGTGCGGG CAGCCTTGTT AGAAAATGCA
                    CATTTTATGG AAAAGTTTTA TCTGCAGAAT GGGAGGTTTA AACAAACATC
               151
55
               201
                    TACCAAGTGG CCAAGTTTGC CGATTAAAGA GGCAGAAGGC TTTTGTATCC
                    GTTTGAATGG AATCGtCGCG CGGG..GCTT TAGACAGTAA ATTCATGTTG
               251
               301
                   AAGGCGGTAG CCATAGATAA AGATAAAAAT CCTTTTATTA TTAAGATGAA
                    TGAAAATCTA GTAACCTTTA ATTTGCAAGA AGTCCGCCAG TTCGTGTAGT
               401
                    GACGGGCTGG ATTATTTTAA AGGAAATGAT AAGGACTGCA AGTTACTTAA
60
               451
                    GTAG
```

5

This corresponds to the amino acid sequence <SEQ ID 820; ORF127>:

```
1 MTDNRGFTLV ELISVVLILS VLALIVYPSY RNYVEKAKIN AVRAALLENA
    HFMEKFYLQN GRFKQTSTKW PSLPIKEAEG FCIRLNGIVA RXALDSKFML
51
    KAVAIDKDKN PFIIKMNENL VTFICKKSAS SCSDGLDYFK GNDKDCKLLK
101
151
```

# Further work revealed the following DNA sequence <SEQ ID 821>:

```
ATGACTGATA ATCGGGGGTT TACGCTGGTT GAATTAATAT CAGTGGTCTT
                    GATATTGTCT GTACTTGCTT TAATTGTTTA TCCGAGCTAT CGCAATTATG
                51
                    TTGAGAAAGC AAAGATAAAT GCAGTGCGGG CAGCCTTGTT AGAAAATGCA
                    CATTTTATGG AAAAGTTTTA TCTGCAGAAT GGGAGGTTTA AACAAACATC
10
                151
                    TACCAAGTGG CCAAGTTTGC CGATTAAAGA GGCAGAAGGC TTTTGTATCC
                201
                    GTTTGAATGG AATCGCGCGC GGGGCTTTAG ACAGTAAATT CATGTTGAAG
                251
                301 GCGGTAGCCA TAGATAAAGA TAAAAATCCT TTTATTATTA AGATGAATGA
                351 AAATCTAGTA ACCTTTATTT GCAAGAAGTC CGCCAGTTCG TGTAGTGACG
                401 GGCTGGATTA TTTTAAAGGA AATGATAAGG ACTGCAAGTT ACTTAAGTAG
15
```

This corresponds to the amino acid sequence <SEQ ID 822; ORF127-1>:

- 1 MTDNRGFTLV ELISVVLILS VLALIVYPSY RNYVEKAKIN AVRAALLENA HFMEKFYLQN GRFKQTSTKW PSLPIKEAEG FCIRLNGIAR GALDSKFMLK
- 101 AVAIDKDKNP FIIKMNENLV TFICKKSASS CSDGLDYFKG NDKDCKLLK\*
- Computer analysis of this amino acid sequence gave the following results: 20

# Homology with a predicted ORF from N. meningitidis (strain A)

ORF127 shows 98.0% identity over a 150aa overlap with an ORF (ORF127a) from strain A of N. meningitidis:

		10	20	30	40	50	60
25	orf127.pep	MTDNRGFTLVELIS	VVLILSVLAL	IVYPSYRNYV	EKAKINAVRA	ALLENAHFMEI	KEYLQN
23	OLLES TO THE	пинини	111111111111		:		KEYLON
	orfl27a	MTDNRGFTLVELIS	VVLILSVLAL 20	30 30	40	50	60
		10	20	30			
30		70	80	90	100	110	120
30	orf127.pep	GRFKQTSTKWPSLE		1111 11 11	111111111	,,,,,,,,,,,,	111111
	orf127a	GRFKQTSTKWPSLI	PIKEAEGFCIR	INGI-ARGAL	DSKFMLKAV	AIDKDKNPFII	KMNENL
		. 70	80	90	100	110	
35		130	140	150			
	orf127.pep	VTFICKKSASSCS	DGLDYFKGNDE	(DCKLLKX			
	orf127a	VTFICKKSASSCS		KDCKLLKX			
40		120 130	140	150			

The complete length ORF127a nucleotide sequence <SEQ ID 823> is:

45 151 CATTTATGG AAAAGTTTTA TCT 201 TACCAAATGG CCAAGTTTGC CGA 251 GTTTGAATGG AATCGCGCGC GGC 301 GCGGTAGCCA TAGATAAAGA TAA 251 AAATCTAGTA ACCTTTATTT GCA	AGTGCGGG CAGCCTTGTT AGAAAATGCA TGCAGAAT GGGAGATTTA AACAAACATC ATTAAAGA GGCAGAAGGC TTTTGTATCC GGCCTTAG ACAGTAAATT CATGTTGAAG AAAATCCT TTTATTATTA AGATGAATGA AAGAAGTC CGCCAGTTCG TGTAGTGACG TGATAAGG ACTGCAAGTT ACTTAAGTAG
---------------------------------------------------------------------------------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

This encodes a protein having amino acid sequence <SEQ ID 824>:

- 1 MTDNRGFTLV ELISVVLILS VLALIVYPSY RNYVEKAKIN TVRAALLENA 51 HFMEKFYLQN GRFKQTSTKW PSLPIKEAEG FCIRLNGIAR GALDSKFMLK
- 101 AVAIDKDKNP FIIKMNENLV TFICKKSASS CSDGLDYFKG NDKDCKLLK\*

60

# ORF127a and ORF127-1 show 99.3% identity in 149 as overlap:

	ORF127a and ORF1	27-1 show 99.3% identity in 149 aa overlap:
5	orf127a.pep orf127-1	10 20 30 40 50 60 MTDNRGFTLVELISVVLILSVLALIVYPSYRNYVEKAKINTVRAALLENAHFMEKFYLON
10	orf127a.pep	70 80 90 100 110 120 GRFKQTSTKWPSLPIKEAEGFCIRLNGIARGALDSKFMLKAVAIDKDKNPFIIKMNENLV
15	orf127a.pep orf127-1	130 140 150 TFICKKSASSCSDGLDYFKGNDKDCKLLKX [                           TFICKKSASSCSDGLDYFKGNDKDCKLLKX 130 140 150
20		edicted ORF from <i>N.gonorrhoeae</i> We identity over a 150 as overlap with a predicted ORF (ORF127ng) from
	N.gonorrhoeae:	
25	orf127.pep	MTDNRGFTLVELISVVLILSVLALIVYPSYRNYVEKAKINAVRAALLENAHFMEKFYLON 60
30	orf127.pep	GRFKQTSTKWPSLPIKEAEGFCIRLNGIVARXALDSKFMLKAVAIDKDKNPFIIKMNENL 120
	orf127.pep orf127ng The complete length	VTFICKKSASSCSDGLDYFKGNDKDCKLLK 150
	The complete length	Old 12/lig illicicotide sequence \SEQ ID 823/ is.
35	51 GAT 101 TTG 151 CAT	ACTGATA ATCGGGGGTT TACACTGGTT GAATTAATAT CAGTGGTCTT CATTGTCT GTACTTGCTT TAATTGTTTA TCCGAGCTAT CGCAATTATG GAGAAAGC AAAGATAAAT GCAGTGCGGG CAGCCTTGTT AGAAAATGCA CTTTATGG AAAAGTTTTA TCTGCAGAAT GGGAGATTTA AACAAACATC CCAAATGG CCAAGTTTGC CGATTAAAGA GGCAGAAGGC TTTTGTATCC
40	251 GTT 301 GCC 351 AAA	TTGAATGG AATCGCGCGC GGGGCTTTAG ACAGTAAATT CATGTTGAAG GGTAGCCA TAGATAAAGA TAAAAATCCT TTTATTATTA AGATGAATGA ATCTAGTA ACCTTTATTT GCAAGAAGTC CGCCAGTTCG TGTAGTGACG CTGGATTA TTTTAAAGGA AATGATAAGG ACTGCAAGTT ACTTAAGTAG
	This encodes a prote	ein having amino acid sequence <seq 826="" id="">:</seq>
45	51 HFN	ONRGFTLV ELISVVLILS VLALIVYPSY RNYVEKAKIN AVRAAFLENA MEKFYLQN GRFKQTSTKW PSLPIKEAEG FCIRLNGIAR GALDSKFMLK AIDKDKNP FIIKMNENLV TFICKKSASS CSDRLDYFKG NDKDCKLLK*
	ORF127ng and ORI	F127-1 show 100.0% identity in 149 aa overlap:
50	orf127-1.pep	10 20 30 40 50 60 MTDNRGFTLVELISVVLILSVLALIVYPSYRNYVEKAKINAVRAALLENAHFMEKFYLQN
	orf127ng-1	MTDNRGFTLVELISVVLILSVLALIVYPSYRNYVEKAKINAVRAALLENAHFMEKFYLQN 10 20 30 40 50 60
55	orf127-1.pep orf127ng-1	70 80 90 100 110 120 GRFKQTSTKWPSLPIKEAEGFCIRLNGIARGALDSKFMLKAVAIDKDKNPFIIKMNENLV
60	· · · · · · · · · · · · · · · · · · ·	70 80 90 100 110 120

	130	140	150
orf127-1.pep	TFICKKSASSCSDG	LDYFKGNDK	DCKLLKX
OLLIZ, T.Pop	111111111111111111111111111111111111111	1111111111	111111
orf127ng-1	TFICKKSASSCSDG	LDYFKGNDK	DCKLLKX
Ollie.mg -	130	140	150

This analysis, including the fact that the predicted transmembrane domain is shared by the meningococcal and gonococcal proteins, suggests that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### 10 Example 98

5

The following partial DNA sequence was identified in N.meningitidis <SEQ ID 827>

```
..GTGTCGCTGG CTTCGGTGAT TGCCTCTCAA ATCTTCCTTT ACGAAGATTT
                     CAACCAAATG CGGAAAACCC GTGGAGCTAT CTGCGGTTTT CTTGTCCAAT
                51
                      ATTTATCTGG GGTTTCAGCA GGGGTATTTC GATTTGAGTG CCGACGAGAA
               101
                      CCCCGTACTG CATATCTGGT CTTTGGCAGT AGAGGAACAG TATTACCTCC
15
               151
                      TGTATCCCCT TTTGCTGATA TTTTGCTGCA AAAAAACCAA ATCGCTACGG
               201
                      GTGCTGCGTA ACATCAGCAT CATCCTGTTT TTGATTTTGA CTGCCTCATC
               251
                      GTTTTTGCCA AGCGGGTTTT ATACCGACAT CCTCAACCAA CCCAATACTT
               301
                      ATTACCTTTC GACACTGAGG TTTCCCGAGC TGTTGGCAGG TTCGCTGCTG
               351
                      GCGGTTTACG GGCAAACGCA AAACGGCAGA CGGCAAACAG CAAATGGAAA
20
               401
                      ACGGCAGTTG CTTTCATCAC TCTGCTTCGG CGCATTGCTT GCCTGCCTGT
               451
                      TCGTGATTGA CAAACACAAT CCGTTTATCC CGGGAATGAC CCTGCTCCTT
               501
                      CCCTGCCTGC TGACGGCACT GCTTATCCGG AGTATGCAAT ACGGGACACT
               551
                      TCCGACCCGC ATCCTGTCGG CAAGCCCCAT CGTATTTGTC GGCAAAATCT
               601
                      25
               651
                      ATTAGAGGCG GGAAACAGCT CGGACTGCCT GCCG..
               701
```

This corresponds to the amino acid sequence <SEQ ID 828; ORF128>:

```
30 ... VSLASVIASQ IFLYEDFNQM RKTVELSAVF LSNIYLGFQQ GYFDLSADEN
51 PVLHIWSLAV EEQYYLLYPL LLIFCCKKTK SLRVLRNISI ILFLILTASS
FLPSGFYTDI LNQPNTYYLS TLRFPELLAG SLLAVYGQTQ NGRRQTANGK
151 RQLLSSLCFG ALLACLFVID KHNPFIPGMT LLLPCLLTAL LIRSMQYGTL
201 PTRILSASPI VFVGKISYSL YLYHWIFIAF APLIRGGKQL GLPA..
```

Further work revealed the complete nucleotide sequence <SEQ ID 829>:

	1	ATGCAAGCTG	TCCGATACAG		G1.0001	GGGCCGTCGC
35	51	CGTGCTATCC	GTCATGATTT	1001001011	IMICCOLLOC	CTGCCCGGAG
33	101	GATTCCTGGG	GGTGGACATT		10101101111	CCTCATTACC
	151	GGCATCATTC	TTTCTGAAAT	*******	101111011	TCCGGGATTT
	201	TTATACCCGC	AGGATTAAGC	GGATTTATCC	1000111111	GCGGCCGTGT
	251	CGCTGGCTTC	GGTGATTGCC	I C I C I M I I C I	100111110011	AGATTTCAAC
40	301	CAAATGCGGA	AAACCGTGGA	GCTTTCTGCG	GTTTTCTTGT	CCAATATTTA
40	351	TCTGGGGTTT	CAGCAGGGGT	ATTTCGATTT	GAGTGCCGAC	GAGAACCCCG
	401	TACTGCATAT	CTGGTCTTTG	GCAGTAGAGG	AACAGTATTA	
	451	CCCCTTTTGC	TGATATTTTG	CTGCAAAAAA	ACCAAATCGC	
	501		AGCATCATCC	TGTTTTTGAT	TTTGACTGCC	TCATCGTTTT
45	551	TGCCAAGCGG	GTTTTATACC	GACATCCTCA	ACCAACCCAA	
77	601	CTTTCGACAC	TGAGGTTTCC	CGAGCTGTTG	GCAGGTTCGC	TGCTGGCGGT
	651	TTACGGGCAA	ACGCAAAACG	GCAGACGGCA	AACAGCAAAT	GGAAAACGGC
	701			TTCGGCGCAT	TGCTTGCCTG	CCTGTTCGTG
	751		ACAATCCGTT	TATCCCGGGA	ATGACCCTGC	TCCTTCCCTG
50	801	CCTGCTGACG	GCACTGCTTA	TCCGGAGTAT		ACACTTCCGA
30	851	CCCGCATCCT	GTCGGCAAGC	CCCATCGTAT	TTGTCGGCAA	AATCTCTTAT
	901	TCCCTATACC	TGTACCATTG	GATTTTTATT	GCTTTCGCCC	ATTACATTAC
	951	AGGCGACAAA		TGCCTGCCGT	ATCGGCGGTT	GCCGCGTTGA
	1001	CGGCCGGATT			TGATTGAACA	GCCGCTTAGA
55	1051	AAACGGAAGA		AAAGGCATTT	TTCTGCCTCT	ATCTCGCCCC
<i>))</i>	1101	GTCCCTGATA			CGCAAGGGGG	ATATTGAAAC
	1151	AGGAACACCT			CCCTTGCTGC	GGAAAATCAT
	1111	MOGNACACCI	000000000			

60

orf128.pep

	1201 m	
	1201 TT 1251 GG	TCCGGAAA CCGTCCTGAC CCTCGGCGAC TCGCACGCCG GACACCTGAG GGTTTCTG GATTATGTCG GCAGCCGGGA AGGGTGGAAA GCCAAAATCC
	1301 TG	TCCCTCGA TTCGGAGTGT TTGGTTTGGG TAGATGAGAA GCCAAAATCC
	1351 AF	CCCGTTAT GTCGAAAATA CCGGGATGAA GTTGAAAAAG CCGAAGCCGT
5	1401 TT	TCATTGCC CAATTCTATG ATTTGAGGAT GGGCGGCCAG CCTGTGCCGA
	1451 GA	NTTTGAAGC GCAATCCTTC CTAATACCCG GGTTCCCAGC CCGATTCAGG
	. 1501 GA	NACCGTCA AAAGGATAGC CGCCGTCAAA CCCGTCTATG TTTTTGCAAA
	1551 C#	ACACATCA ATCAGCCGTT CGCCCCTGAG GGAGGAAAAA TTGAAAAGAT
10	1601 TT	GCCGCAAA CCAATATCTC CGCCCCATTC AGGCTATGGG CGACATCGGC
10	1651 AF	AGAGCAATC AGGCGGTCTT TGATTTGATT AAAGATATTC CCAATGTGCA
	1701 TT 1751 GC	GGGTGGAC GCACAAAAAT ACCTGCCCAA AAACACGGTC GAAATATACG
	1751 GC 1801 TA	CGCTATCT TTACGGCGAC CAAGACCACC TGACCTATTT CGGTTCTTAT ATATGGGGC GGGAATTCCA CAAACACGAA CGCCTGCTTA AATCTTCCCA
	1851 · CG	GCGGCGCA TTGCAGTAG
	-	•
15	This corresponds to	the amino acid sequence <seq 830;="" id="" orf128-1="">:</seq>
	•	
	1 MC	AVRYRPEI DGLRAVAVLS VMIFHLNNRW LPGGFLGVDI FFVISGFLIT
	51 <u>GI</u>	ILSEIQNG SFSFRDFYTR RIKRIYPAFI AAVSLASVIA SQIFLYEDFN
	101 QN 151 PI	RKTVELSA VFLSNIYLGF QQGYFDLSAD ENPVLHIWSL AVEEQYYLLY LLIFCCKK TKSLRVLRNI SIILFLILTA SSFLPSGFYT DILNQPNTYY
20	201 LS	TERFPELL AGSELAVYGQ TONGRROTAN GKROLESSIC FGALLACLFV
20	251 II	OKHNPFIPG MTLLLPCLLT ALLIRSMQYG TLPTRILSAS PIVFVGKISY
		YLYHWIFI AFAHYITGDK QLGLPAVSAV AALTAGFSLL SYYLIEQPLR
	351 KF	KMTFKKAF FCLYLAPSLI LVGYNLYARG ILKQEHLRPL PGAPLAAENH
	401 FE	PETVLTLGD SHAGHLRGFL DYVGSREGWK AKILSLDSEC LVWVDEKLAD
25	451 NE	PLCRKYRDE VEKAEAVFIA QFYDLRMGGQ PVPRFEAQSF LIPGFPARFR
	501 ET	YKRIAAVK PVYVFANNTS ISRSPLREEK LKRFAANOYL RPIOAMGDIG
	551 KS	NQAVFDLI KDIPNVHWVD AQKYLPKNTV EIYGRYLYGD QDHLTYFGSY
	601 YN	GREFHKHE RLLKSSHGGA LQ*
	Computer analysis	of this amino acid sequence gave the following results:
30	Homology with hypor	thetical integral membrane protein HI0392 of H.influenzae (accession number U32723)
	ORF128 and H103	92 show 52% aa identity in 180aa overlap:
	Orf128: 1	VSLASVIASQIFLYEDFNQMRKTVELSAVFLSNIYLGFQQGYFDLSADENPVLHIWSLAV 60
		++L S IAS IF+Y DFN++RKT+EL+ FLSN YLG OGYFDLSA+ENPVLHIWSLAV
	HI0392: 46	MALVSFIASAIFIYNDFNKLRKTIELAIAFLSNFYLGLTQGYFDLSANENPVLHIWSLAV 105
35		
	Orf128: 61	EEQXXXXXXXXIFCCKKTKSLRVLRNISIILFLILTASSFLPSGFYTDILNQPNTYYLS 120
	HT0302: 106	E Q I KK + ++VL I++ILF IL A+SF+ + FY ++L+QPN YYLS EGQYYLIFPLILILAYKKFREVKVLFIITLILFFILLATSFVSANFYKEVLHQPNIYYLS 165
	n10392. 100	EGGIIPILEPIPIPULVELVEAVAPETITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTETTELITETTELITETTELITETTELITETTELITETTELITETTELITETTE
40	Orf128: 121	TLRFPELLAGSLLAVYGQTQNGRRQTANGKRQLLSSLCFGALLACLFVIDKHNPFIPGMT 180
		LRFPELL GSLLA+Y N + Q + +L+ L L +CLF+++ + FIPG+T
	HI0392: 166	NLRFPELLVGSLLAIYHNLSN-KVQLSKQVNNILAILSTLLLFSCLFLMNNNIAFIPGIT 224
	II am al a an mith a m	andieted ODE Com Mr. 1 1/11/11 (4.1.4)
	Homology with a p	oredicted ORF from N. meningitidis (strain A)
45	ORF128 shows 98	0% identity over a 244aa overlap with an ORF (ORF128a) from strain A of N.
73	Old 120 5110 WB 70.	over deficitly over a 244aa overlap with all ORF (ORF 126a) from strain A 01 14.
	meningitidis:	
		10 20 30
	orf128.pep	VSLASVIASQIFLYEDFNQMRKTVELSAVF
50	- m 61 00 -	111111111111111111111111111111111111111
30	orf128a	ILSEIQNGSFSFRDFYTRRIKRIYPAFIAAVSLASVIASQIFLYEDFNQMRKTVELSAVF
	,	60 70 80 90 100 110
		40 50 60 70 80 90
= :	orf128.pep	LSNIYLGFQQGYFDLSADENPVLHIWSLAVEEQYYLLYPLLLIFCCKKTKSLRVLRNISI
55		
	orf128a	LSNIYLGFQQGYFDLSADENPVLHIWSLAVEEQYYLLYPLLLIFCCKKTKSLRVLRNISI
		120 130 140 150 160 170

100 110 120 130 140 150 ILFLILTASSFLPSGFYTDILNQPNTYYLSTLRFPELLAGSLLAVYGQTQNGRRQTANGK

	orf128a	ILFLILTATSFLPSGFYTDILNQPNTYYLSTLRFPELLAGSLLAVYGQTQNGRRQTANGK 180 190 200 210 220 230
		160 170 180 190 200 210 RQLLSSLCFGALLACLFVIDKHNPFIPGMTLLLPCLLTALLIRSMQYGTLPTRILSASPI
5	orf128.pep	
	orf128a	RQLLSSLCFGALLACLFVIDKHNPFIPGMTLLLPCLLTALLIRSMQYGTLPTRILSASPI 240 250 260 270 280 290
10	orf128.pep	220 230 240 VFVGKISYSLYLYHWIFIAFAPLIRGGKQLGLPA
	orf128a	VFVGKISYSLYLYHWIFIAFAHYITGDKQLGLPAVSAVAALTAGFSLLSYYLIEQPLRKR
15	orf128a	XMTFKKAFFCLYLAPSLILVGYNLYARGILKQEHLRPLPGAPLAAENHFPETVLTLGDSH 360 370 380 390 400 410
	The complete leng	gth ORF128a nucleotide sequence <seq 831="" id=""> is:</seq>
	=	ATGCAAGCTG TCCGATACAG ACCGGAAATT GACGGATTGC GGGCCGTCGC
20		CONCOUNTION CHICAMENTO TOCACCTERA TARCCECTES CISCOCESAS
20		GATTCTTGGG GGTGGACATT TTCTTTGTCA TCTCAGGATT CCTCATTACC GGCATCATTC TTTCTGAAAT ACAGAACGGT TCTTTTTCTT TCCGGGATTT
	001	TO THE TAKE THE PROPERTY AND A CONTRACT CONTRACT OF THE PROPERTY OF THE PROPER
		COMOCOMBC COMCAMMCCC MCMCAAAMCT TCCTTTACGA AGAIIICAAC
25		CAAATGCGGA AAACCGTGGA GCTTTCTGCG GTTTTCTTGT CCAATATTTA TCTGGGGTTT CAGCAGGGGT ATTTCGATTT GAGTGCCGAC GAGAACCCCG TCTGGGGTTT CAGCAGGGGT ATTTCGATTT GAGTGCCGAC GAGAACCCCG
		MACHOCAMAM CMCCMCMMMC GCAGMAGAGG AACAGTATIA CCICCIGIAI
		COMCOMMENCO MCAMAMPROTO CTCCAAAAAA ACAAAATCGC TACGGGIGCI
	r 0.1	COCHANCARC ACCARCARCE TAPPTCTGAT TTTGACTGCC ACAICGIIII
30	551	TGCCAAGCGG GTTTTATACC GATATTCTCA ACCAACCCAA TACTTATTAC CTTTCGACAC TGAGGTTTCC CGAGCTGTTG GCAGGTTCGC TGCTGGCGGT
		CTTTCGACAC TGAGGTTTCC CGAGCTGTTG GCAGACAAAT GGAAAACGGC TTACGGGCAA ACGCAAAACG GCAGACGGCA AACAGCAAAT GGAAAACGGC
	651 701	TOWNSON THE TOWN THE THE THE THE THE THE TOWN TH
	751	AMECACAAAC ACAATCCGTT TATCCCGGGA ATGACCCTGC TCCTTCCCTG
35	801	CONCORCA CO GCACTGCTTA TCCGGAGTAT GCAATACGGG ACACTICCGA
•	851	CCCGCATCCT GTCGGCAAGC CCCATCGTAT TTGTCGGCAA AATCTCTTAT TCCCTATACC TGTACCATTG GATTTTATT GCTTTCGCCC ATTACATTAC
	901	PROCERGY AND CACCERGEAR TECCEFFICET ATCGGCGGTT GCCGCGTTGA
	951 1001	COCCOCAME EMCCCOMOTO ACTATATA TGATTGAACA GCCGCIIAGA
40	1051	TARGOCANCA WCACCOWCAA AAACGCATTT TICIGCCICI AICICGCCC
40	1101	ORGONICAMA CHMCMCCCCMM ACAACCMGTA CGCAAGGGGG AIAIIGAAAC
	1151	AGGAACACCT CCGCCCGTTG CCCGGCGCG CCCTTGCTGC GGAAAATCAT TTTCCGGAAA CCGTCCTGAC CCTCGGCGAC TCGCACGCCG GACACCTGCG
	1201	ACCOMPROME CAMPATEREE ECACCECCA AUGULTUGAAA UUUAAAA
45	1251 1301	momocomoca mmcccacemen precepting GGG TAGATGAGAA GCIGGCAGAC
73	1351	AACCCCMMAM CMCCAAAAAA CCGGGATGAA GTTGAAAAAG CCGAAGCCGI
	1401	mmmcammccc caammcmamc ammigaccar Guguuduu ceesiseedii
	1451	GATTTGAAGC GCAATCCTTC CTAATACCCG GGTTCCCAGC CCGATTCAGG GAAACCGTCA AAAGGATAGC CGCCGTCAAA CCCGTCTATG TTTTTGCAAA
50	1501 1551	CARCACACA ATCACCCTT CCCCCTGAG GGAGGAAAA TIGAAAAAI
30	1601	THE COCCARA CORRESPONDE COCCCCCATTO AGGCTATGGG CGACALCGGC
	1651	ARCACCAARC ACCCCCTCTT TCATTTGATT AAAGATATTC CCAAIGIGCA
	1701	
<i>E E</i>	1751 1801	
55	1851	CGACGGCGCA TTGCAGTAG
		the leaving amino said sequence <sfo 832="" id="">:</sfo>
	This encodes a	protein having amino acid sequence <seq 832="" id="">:</seq>
	1	TENERS DESCRIPTION DESCRIPTION DE LA SOLICIO DE N
<i>c</i> n	51 101	ONDUME I ON VET SNIVI OF OOGYFDI SAD ENPVLHIMSL AVECUITED
60	151	DITITECTV TYCIDVIRNT STILFILITA TSPLYSGEIT DIENQUALIT
	201	TOWN DEDELT ACCULATION TONGRESTAN GEROLLSSLC FGALLACHEV
	251	TOWNSHIP MILITORITY ALLTRIMOYG TEPTRILISAS PIVIVGALSI
	301	
65	351	EDERGIA MICH CHACHLREFT, DYVESREGWK AKILSLUSEC DYWYDERLAND
	401 451	
	42T	THE MODIFICATION CONTRACTOR OF THE PROPERTY OF

- 501 ETVKRIAAVK PVYVFANNTS ISRSPLREEK LKRFAANQYL RPIQAMGDIG 551 KSNQAVFDLI KDIPNVHWVD AQKYLPKNTV EIYGRYLYGD QDHLTYFGSY 601 YMGREFHKHE RLLKSSRDGA LQ\*

## ORF128a and ORF128-1 show 99.5% identity in 622 aa overlap:

5	orf128a.pep	MQAVRYRPEIDGLRAVAVLSVMIFHLNNRWLPGGFLGVDIFFVISGFLITGIILSEIQNG
	orf128-1	MQAVRYRPEIDGLRAVAVLSVMIFHLNNRWLPGGFLGVDIFFVISGFLITGIILSEIQNG
10	orf128a.pep	SFSFRDFYTRRIKRIYPAFIAAVSLASVIASQIFLYEDFNQMRKTVELSAVFLSNIYLGF
	orf128-1	SFSFRDFYTRRIKRIYPAFIAAVSLASVIASQIFLYEDFNQMRKTVELSAVFLSNIYLGF
	orf128a.pep	QQGYFDLSADENPVLHIWSLAVEEQYYLLYPLLLIFCCKKTKSLRVLRNISIILFLILTA
15	orf128-1	QQGYFDLSADENPVLHIWSLAVEEQYYLLYPLLLIFCCKKTKSLRVLRNISIILFLILTA
	orf128a.pep	TSFLPSGFYTDILNQPNTYYLSTLRFPELLAGSLLAVYGQTQNGRRQTANGKRQLLSSLC:
20	orf128-1	SSFLPSGFYTDILNQPNTYYLSTLRFPELLAGSLLAVYGQTQNGRRQTANGKRQLLSSLC
	orf128a.pep	FGALLACLFVIDKHNPFIPGMTLLLPCLLTALLIRSMQYGTLPTRILSASPIVFVGKISY
	orf128-1	FGALLACLFVIDKHNPFIPGMTLLLPCLLTALLIRSMQYGTLPTRILSASPIVFVGKISY
25	orf128a.pep	SLYLYHWIFIAFAHYITGDKQLGLPAVSAVAALTAGFSLLSYYLIEQPLRKRKMTFKKAF
	orf128-1	SLYLYHWIFIAFAHYITGDKQLGLPAVSAVAALTAGFSLLSYYLIEQPLRKRKMTFKKAF
30	orf128a.pep	FCLYLAPSLILVGYNLYARGILKQEHLRPLPGAPLAAENHFPETVLTLGDSHAGHLRGFL
÷	orf128-1	FCLYLAPSLILVGYNLYARGILKQEHLRPLPGAPLAAENHFPETVLTLGDSHAGHLRGFL
	orf128a.pep	DYVGSREGWKAKILSLDSECLVWVDEKLADNPLCRKYRDEVEKAEAVFIAQFYDLRMGGQ
35	orf128-1	DYVGSREGWKAKILSLDSECLVWVDEKLADNPLCRKYRDEVEKAEAVFIAQFYDLRMGGQ
	orf128a.pep	PVPRFEAQSFLIPGFPARFRETVKRIAAVKPVYVFANNTSISRSPLREEKLKRFAANQYL
40	orf128-1	PVPRFEAQSFLIPGFPARFRETVKRIAAVKPVYVFANNTSISRSPLREEKLKRFAANQYL
	orf128a.pep	RPIQAMGDIGKSNQAVFDLIKDIPNVHWVDAQKYLPKNTVEIYGRYLYGDQDHLTYFGSY
45	orf128-1	RPIQAMGDIGKSNQAVFDLIKDIPNVHWVDAQKYLPKNTVEIYGRYLYGDQDHLTYFGSY
45	orf128a.pep	YMGREFHKHERLLKSSRDGALQX
	orf128-1	YMGREFHKHERLLKSSHGGALQX

### Homology with a predicted ORF from N.gonorrhoeae

ORF128 shows 93.4% identity over 244 as overlap with a predicted ORF (ORF128ng) from N. gonorrhoeae:

٠.	orf128.pep	VSLASVIASQIFLYEDFNQMRKTVELSAVF	30
55	orf128ng		112
	orf128.pep	LSNIYLGFQQGYFDLSADENPVLHIWSLAVEEQYYLLYPLLLIFCCKKTKSLRVLRNISI	90
	orf128ng	LSNIYLGFRLGYFDLSADENPVLHIWSLAVEEQYYLLYPLLLIFCYKKTKSLRVLRNISI	172
60	orf128.pep	ILFLILTASSFLPSGFYTDILNQPNTYYLSTLRFPELLAGSLLAVYGQTQNGRRQTANGK	150
	orf128ng		232
65	orf128.pep	RQLLSSLCFGALLACLFVIDKHNPFIPGMTLLLPCLLTALLIRSMQYGTLPTRILSASPI	210
	orf128ng		292

```
244
                       VFVGKISYSLYLYHWIFIAFAPLIRGGKQLGLPA
          orf128.pep
                       111111111111111111111
                                             _1 | | | | | | | | | | | | | | |
                       VFVGKISYSLYLYHWIFIAFAHYITGDKQLGLPAVSAVAALTAGFSLLSYYLIEQPLRKR
                                                                                   352
          orf128ng
     The complete length ORF128ng nucleotide sequence <SEQ ID 833> is:
5
                    ATGCAAGCTG TCCGATACAG GCCTGAAATT GACGGATTGC GGGCCGTCGC
                    CGTGCTATCC GTCATTATTT TCCACCTGAA TAACCGCTGG CTGCCCGGAG
                51
                    GATTCCTGGG GGTGGACATT TTCTTTGTCA TCTCGGGATT CCTCATTACC
               101
                    AACATCATTC TTTCTGAAAT ACAGAACGGT TCTTTTTCTT TCCGGGATTT
               151
                    TTATACCCGC AGGATTAAGC GGATTTATCC TGCTTTTATT GCGGCCGTGT
10
               201
                    CCCTGGCTTC GGTGATTGCT TCTCAAATCT TCCTTTACGA AGATTTCAAC
               251
                    CAAATGAGGA AAACCATAGA GCTTTCTACG GTTTTTTTGT CCAATATTTA
               301
                    TTTGGGGTTC CGATTGGGGT ATTTCGATTT GAGTGCCGAC GAGAACCCCG
               351
                    TACTGCATAT CTGGTCTTTG GCGGTAGAGG AACAGTATTA CCTCCTGTAT
               401
                    CCTCTTTTGC TGATATTCTG TTACAAAAAA ACCAAATCAC TACGGGTGCT
15
               451
                    GCGTAATATC AGCATCATCC TGTTTCTGAT TTTGACCGCA TCATCGTTTT
                501
                     TGCCGGCCGG GTTTTATACC GACATCCTCA ACCAACCCaa TACTTATTAC
                551
                     CTTTCGACAC TGAGGTTTCC CGAGCTGTTG GTGGGTTCGC TGTTGGCGGT
                601
                     TTACGGGCAA ACGCAAAACG GCAGACGGCA AACAGAAAAT GGAAAACGGC
                651
                     AGTTGCTTTC ATTACTCTGT TTCGGCGCat tgCTTGTCTG CCTGTTCGTG
                701
20
                     ATCGACAAAC ACGATCCGTT TATCCCGGGA ATAACCCTGC TCCTTCCCTG
                751
                     CCTGCTGACG GCGCTGCTTA TCCGGAGTAT GCAATACGGG ACACTTCCGA
                801
                     CCCGCATCCT GTCGGCAAGC CCCATCGTAT TTGTCGGCAA AATCTCTTAT
                851
                     901
                     AGGCGACAAA CAGCTCGGAC TGCCTGCCGT ATCGGCGGTT GCCGCGTTGA
25
                951
                     CGGCCGGATT TTCCCTGTTG AGCTATTATT TGATTGAACA GCCGCTTAGA
               1001
                     AAACGGAAGA TGACCTTCAA AAAGGCATTT TTCTGCCTTT ATCTCGCCCC
               1051
                     GTCCCTGATG CTTGTCGGTT ACAACCTGTA TTCAAGAGGG ATATTGAAAC
               1101
                     AGGAACACCT CCGCCCGCTG CCCGGCACGC CCGTTGCTGC GGAAAATAAT
               1151
                     TTTCCGGAAA CCGTCTTGAC CCTCGGCGAC TCGCACGCCG GACACCTGCG
 30
               1201
                     GGGGTTTCTG GATTATGTCG GCGGCAGGGA AGGGTGGAAA GCTAAAATCC
               1251
                     TGTCCCTCGA TTCGGAGTGT TTGGTTTGGG TGGATGAGAA GCTGGCAGAC
               1301
                     AACCCGTTGT GCCGAAAATA CCGGGATGAA GTTGAAAAAG CCGAAGCTGT
               1351
                     TTTCATTGCC CAATTCTATG ATTTGAGGAT GGGCGGCCAG CCCGTGCCGA
               1401
                     GATTTGAAGC GCAATCCTTC CTGATACCCG GGTTCAAAGC CCGATTCAGG
 35
               1451
                     GAAACCGTCA AGAGGATAGC CGCCGTCAAA CCTGTATATG TTTTTGCAAA
               1501
                     CAATACATCA ATCAGCCGTT CTCCCTTGAG GGAGGAAAAA TTGAAAAGAT
                1551
                     TTGCTATAAA CCAATACCTC CGGCCTATTC GGGCTATGGG CGACATCGGC
                1601
                     AAGAGCAATC AGGCGGTCTT TGATTTGGTT AAAGATATTC CCAATGTGCA
                1651
                      TTGGGTGGAC GCACAAAAAT ACCTGCCCAA AAACACGGTC GAAATACACG
                1701
 40
                      GACGCTATCT TTACGGCGAC CAAGACCACC TGACCTATTT CGGTTCTTAT
                1751
                      TATATGGGGC GGGAATTTCA CAAACACGAA CGCCTGCTCA AGCATTCCCG
                1801
                      AGGCGGCGCA TTGCAGTAG
       This encodes a protein having amino acid sequence <SEQ ID 834>:
                      MQAVRYRPEI DGLRAVAVLS VIIFHLNNRW LPGGFLGVDI FFVISGFLIT
  45
                      NIILSEIONG SFSFRDFYTR RIKRIYPAFI AAVSLASVIA SQIFLYEDFN
                      QMRKTIELST VFLSNIYLGF RLGYFDLSAD ENPVLHIWSL AVEEQYYLLY
                 151 PLLLIFCYKK TKSLRVLRNI SIILFLILTA SSFLPAGFYT DILNQPNTYY
                     LSTLRFPELL VGSLLAVYGQ TONGRROTEN GKROLLSLLC FGALLVCLFV
                      IDKHOPFIPG ITLLLPCLLT ALLIRSMOYG TLPTRILSAS PIVFVGKISY
  50
                 251
                      SLYLYHWIFI AFAHYITGDK QLGLPAVSAV AALTAGFSLL SYYLIEQPLR
                      KRKMTFKKAF FCLYLAPSLM LVGYNLYSRG ILKQEHLRPL PGTPVAAENN
                 351
                      FPETVLTLGD SHAGHLRGFL DYVGGREGWK AKILSLDSEC LVWVDEKLAD
                      NPLCRKYRDE VEKAEAVFIA QFYDLRMGGQ PVPRFEAQSF LIPGFKARFR
                 451
                      ETVKRIAAVK PVYVFANNTS ISRSPLREEK LKRFAINQYL RPIRAMGDIG
  55
                      KSNQAVFDLV KDIPNVHWVD AQKYLPKNTV EIHGRYLYGD QDHLTYFGSY
                      YMGREFHKHE RLLKHSRGGA LQ*
       ORF128ng and ORF128-1 show 95.7% identity in 622 aa overlap:
                           {\tt MQAVRYRPEIDGLRAVAVLSVMIFHLNNRWLPGGFLGVDIFFVISGFLITGIILSEIQNG}
             orf128-1.pep
                            ининининин: попинининий пирин: пини
  60
                            MQAVRYRPEIDGLRAVAVLSVIIFHLNNRWLPGGFLGVDIFFVISGFLITNIILSEIQNG
             orf128ng
                            SFSFRDFYTRRIKRIYPAFIAAVSLASVIASQIFLYEDFNQMRKTVELSAVFLSNIYLGF
             orf128-1.pep
                            SFSFRDFYTRRIKRIYPAFIAAVSLASVIASQIFLYEDFNQMRKTIELSTVFLSNIYLGF
   65
             orf128ng
```

	orf128-1.pep	QQGYFDLSADENPVLHIWSLAVEEQYYLLYPLLLIFCCKKTKSLRVLRNISIILFLILTA :
5	orf128ng	RLGYFDLSADENPVLHIWSLAVEEQYYLLYPLLLIFCYKKTKSLRVLRNISIILFLILTA
3	orf128-1.pep	SSFLPSGFYTDILNQPNTYYLSTLRFPELLAGSLLAVYGQTQNGRRQTANGKRQLLSSLC
	orf128ng	SSFLPAGFYTDILNQPNTYYLSTLRFPELLVGSLLAVYGQTQNGRRQTENGKRQLLSLLC
10	orf128-1.pep	FGALLACLFVIDKHNPFIPGMTLLLPCLLTALLIRSMQYGTLPTRILSASPIVFVGKISY
	orf128ng	FGALLVCLFVIDKHDPFIPGITLLLPCLLTALLIRSMQYGTLPTRILSASPIVFVGKISY
15	orf128-1.pep	SLYLYHWIFIAFAHYITGDKQLGLPAVSAVAALTAGFSLLSYYLIEQPLRKRKMTFKKAF
	orf128ng	SLYLYHWIFIAFAHYITGDKQLGLPAVSAVAALTAGFSLLSYYLIEQPLRKRKMTFKKAF
	orf128-1.pep	FCLYLAPSLILVGYNLYARGILKQEHLRPLPGAPLAAENHFPETVLTLGDSHAGHLRGFL
20	orf128ng	FCLYLAPSLMLVGYNLYSRGILKQEHLRPLPGTPVAAENNFPETVLTLGDSHAGHLRGFL
	orf128-1.pep	DYVGSREGWKAKILSLDSECLVWVDEKLADNPLCRKYRDEVEKAEAVFIAQFYDLRMGGQ
25	orf128ng	DYVGGREGWKAKILSLDSECLVWVDEKLADNPLCRKYRDEVEKAEAVFIAQFYDLRMGGQ
	orf128-1.pep	PVPRFEAQSFLIPGFPARFRETVKRIAAVKPVYVFANNTSISRSPLREEKLKRFAANQYL
	orf128ng	PVPRFEAQSFLIPGFKARFRETVKRIAAVKPVYVFANNTSISRSPLREEKLKRFAINQYL
30	orf128-1.pep	RPIQAMGDIGKSNQAVFDLIKDIPNVHWVDAQKYLPKNTVEIYGRYLYGDQDHLTYFGSY
•	orf128ng	RPIRAMGDIGKSNQAVFDLVKDIPNVHWVDAQKYLPKNTVEIHGRYLYGDQDHLTYFGSY
35	orf128-1.pep	YMGREFHKHERLLKSSHGGALQX
	orf128ng	YMGREFHKHERLLKHSRGGALQX 610 620

In addition, ORF218ng shows homology to a hypothetical *H.influenzae* protein:

```
sp|P43993|Y392 HAEIN
                                 HYPOTHETICAL
                                                PROTEIN
                                                          HI0392
                                                                    >gi|1074385|pir||B64007
40
          hypothetical protein HI0392 - Haemophilus influenzae (strain Rd KW20)
          >gi|1573364 (U32723) H. influenzae predicted coding region HI0392 [Haemophilus
          influenzae] Length = 245
           Score = 239 bits (604), Expect = 3e-62
           Identities = 124/225 (55%), Positives = 152/225 (67%), Gaps = 1/225 (0%)
45
          Query: 38 VDIFFVISGFLITNIILSEIQNGSFSFRDFYTRRIKRIYPXXXXXXXXXXXXXXXXXFLYE 97
                     +DIFFVISGFLIT II++EIQ SFS + FYTRRIKRIYP
                     MDIFFVISGFLITGIIITEIQQNSFSLKQFYTRRIKRIYPAFITVMALVSFIASAIFIYN 60
          Sbjct: 1
50
                     DFNQMRKTIELSTVFLSNIYLGFRLGYFDLSADENPVLHIWSLAVEEQXXXXXXXXXIFC 157
          Query: 98
                     DFN++RKTIEL+ FLSN YLG GYFDLSA+ENPVLHIWSLAVE Q
          Sbjct: 61 DFNKLRKTIELAIAFLSNFYLGLTQGYFDLSANENPVLHIWSLAVEGQYYLIFPLILILA 120
          Query: 158 YKKTKSLRVLRNISIILFLILTASSFLPAGFYTDILNQPNTYYLSTLRFPELLVGSLLAV 217
55
                     YKK + ++VL I++ILF IL A+SF+ A FY ++L+QPN YYLS LRFPELLVGSLLA+
          Sbjct: 121 YKKFREVKVLFIITLILFFILLATSFVSANFYKEVLHQPNIYYLSNLRFPELLVGSLLAI 180
          Query: 218 YGQTQNGRRQTENGKRQLLSLLCFGALLVCLFVIDKHDPFIPGIT 262
                          N + Q
                                      +L++L
                                               L CLF+++ + FIPGIT
60
           Sbjct: 181 YHNLSN-KVQLSKQVNNILAILSTLLLFSCLFLMNNNIAFIPGIT 224
```

This analysis, including the identification of several putative transmembrane domains, suggests that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

## Example 99

The following partial DNA sequence was identified in N.meningitidis <SEQ ID 835>:

```
1 ..ATTATTTACG AATACCGCTG GATGTTTCTT TACGGCGCAC TGACGACCTT

51 GGGGCTGACG GTCGTGGCAA C..GCGGGCGG TTCGGTATTG GGTCTGTTGT

51 TGGCGTTGGC GCGCCTGATT CACTTGGAAA AAGCCGGTGC GCCGATGCGC

151 GTGCTGGCGT GGGCGTTGCG TAAAGTTTCG CTGCTGTATG TTACGCTGTT

201 CCGGGGTACG CCGCTGTTTG TGCAGATTGT GATTTGGGCG TATGTGTGGT

251 TTCCGTTTTT CGTC..
```

This corresponds to the amino acid sequence <SEQ ID 836; ORF129>:

1 ...IIYEYRWMFL YGALTTLGLT VVAXAGGSVL GLLLALARLI HLEKAGAPMR
51 VLAWALRKVS LLYVTLFRGT PLFVQIVIWA YVWFPFFV..

Further work revealed the complete nucleotide sequence <SEQ ID 837>:

15	1 51 101 151 201 251 301	CGGCGCACTG CGGTATTGGG GCCGGTGCGC GCTGTATGTT TTTGGGCGTA	ACGACCTTGG TCTGTTGTTG CGATGCGCGT ACGCTGTTCC TGTGTGGTTT GCGAGGCGGC	TATTTACGAA GGCTGACGGT GCGTTGGCGC GCTGGCGTGG GGGGTACGCC CCGTTTTTCG AATCGCGCTG	CGTGGCAACG GCCTGATTCA GCGTTGCGTA GCTGTTTGTG TCCATCCTTC CGTCGCGGAT	GCGGGCGGTT CTTGGAAAAA AAGTTTCGCT CAGATTGTGA AGACGGCATT ACGGGCCGCT
20	351 401 451	GATTGCCGGT AGATTTTCCG GCGCGTTCTT	TCTTTGGCAC CGCGGGCATC TGGGGCTGAC	TGATCGCCAA CAGTCTATAG CTATCCGCAG	CTCGGGGGCG ACAAAGGACA GCGATGCGCT	GATGGAGGCG ATGTGATTCT
25	501 551 601 651 701	CGCTCTTGAA GCGTATGTTC	AGACAGCTCG AGAATACGAT GTCGCCCTGA	TGCTGCCGCC CTGCTGTCGG TACGGGCCGG TTTATCTGTT AAACGTTACA	TCATTGCTGT TATTCGGTTT GATGACGACT	GGCGGAGTTG ATGAAGAACC TTCTTAGGCT

This corresponds to the amino acid sequence <SEQ ID 838; ORF129-1>:

```
1 MDFRFDIIYE YRWMFLYGAL TTLGLTVVAT AGGSVLGLLL ALARLIHLEK
30 51 AGAPMRVLAW ALRKVSLLYV TLFRGTPLFV QIVIWAYVWF PFFVHPSDGI
101 LVSGEAAIAL RRGYGPLIAG SLALIANSGA YICEIFRAGI QSIDKGQMEA
151 ARSLGLTYPQ AMRYVILPQA LRRMLPPLAS EFITLLKDSS LLSVIAVAEL
201 AYVQNTITGR YSVYEEPLYT VALIYLLMTT FLGWIFLRLE KRYNPQHR*
```

Computer analysis of this amino acid sequence gave the following results:

# 35 Homology with a predicted ORF from N. meningitidis (strain A)

ORF129 shows 98.9% identity over a 88aa overlap with an ORF (ORF129a) from strain A of N. meningitidis:

40	orf129.pep	1 IIYEYRWMF          MDFRFDIIYEYRWMF 10	111111111	LTVVAXAGGS	1111111111	40 RLIHLEKAGAE            RLIHLEKAGAE 50	
45	orf129.pep orf129a	60 7 ALRKVSLLYVTLFRG				GEAAIALRRG 110	YGP <u>LIAG</u> 120
50	orf129a	SLALIANSGAYICE 130	FRAGIQSI 140	DKG <u>O</u> MEAARSI 150	LGLTYPQAMR 160	YVILPQALRR 170	MLPPLAS 180

The complete length ORF129a nucleotide sequence <SEQ ID 839> is:

<sup>1</sup> ATGGATTTC GTTTTGACAT TATTTACGAA TACCGCTGGA TGTTTCTTTA

<sup>51</sup> CGGCGCACTG ACGACCTTGG GGCTGACGGT CGTGGCGACG GCGGGCGGTT

		·	
		GTATTGGG TCTGTTGTTG GCGTTGGCGC GCCTGATTCA CTTGGAAAAA	
		CGGTGCGC CGATGCGCGT GCTGGCGTGG GCGTTGCGTA AGGTTTCGCT	
		TGTATGTT ACGCTGTTCC GGGGTACGCC GCTGTTTGTG CAGATTGTGA	
_		TGGGCGTA TGTGTGGTTT CCGTTTTTCG TCCATCCTTC AGACGGCATT	
5		GGTTAGCG GCGAGGCGGC AATCGCGCTG CGTCGCGGAT ACGGGCCGCT	
		TTGCCGGT TCTTTGGCAC TGATCGCCAA CTCGGGGGCG TATATCTGTG	
		ATTTTCCG CGCGGGCATC CAGTCTATAG ACAAAGGACA GATGGAGGCG	
		GCGTTCTT TGGGGCTGAC CTATCCGCAG GCGATGCGCT ATGTGATTCT	
10	501 GC	CGCAGGCA TTGCGCCGTA TGCTGCCGCC TTTGGCGAGC GAGTTCATCA	
10		CTCTTGAA AGACAGCTCG CTGCTGTCGG TCATTGCTGT GGCGGAGTTG	
		GTATGTTC AGAATACGAT TACGGGCCGG TATTCGGTTT ATGAAGAACC	
		TTTACACC GTCGCCCTGA TTTATCTGTT GATGACGACT TTCTTAGGCT ATATTCCT GCGTTTGGAA AAACGTTACA ATCCGCAACA CCGCTGA	
	701 GG	ATATICCT GCGTTTGGAR AMACGITACA ATCCGCAACA CCGCTGA	
	This encodes a prot	ein having amino acid sequence <seq 840="" id="">:</seq>	
15	1 MD	FRFDIIYE YRWMFLYGAL TTLGLTVVAT AGGSVLGLLL ALARLIHLEK	
13		APMRVLAW ALRKVSLLYV TLFRGTPLFV OIVIWAYVWF PFFVHPSDGI	
		SGEAAIAL RRGYGPLIAG SLALIANSGA YICEIFRAGI QSIDKGOMEA	
		SLGLTYPO AMRYVILPOA LRRMLPPLAS EFITLLKDSS LLSVIAVAEL	
		VONTITGR YSVYEEPLYT VALIYLIMTT FLGWIFLRLE KRYNPOHR*	
20	ORF129a and ORF	129-1 show 100.0% identity in 248 aa overlap:	
		·	
	orf129a.pep	MDFRFDIIYEYRWMFLYGALTTLGLTVVATAGGSVLGLLLALARLIHLEKAGAPMRVLAW	
	orf129-1	MDED EDIT VENUE VOLUME VOLUME OF THE STATE O	
	011159-1	MDFRFDIIYEYRWMFLYGALTTLGLTVVATAGGSVLGLLLALARLIHLEKAGAPMRVLAW	
25	orf129a.pep	ALRKVSLLYVTLFRGTPLFVQIVIWAYVWFPFFVHPSDGILVSGEAAIALRRGYGPLIAG	
	orf129-1	ALRKVSLLYVTLFRGTPLFVQIVIWAYVWFPFFVHPSDGILVSGEAAIALRRGYGPLIAG	
οĠ	orf129a.pep	SLALIANSGAYICEIFRAGIQSIDKGQMEAARSLGLTYPQAMRYVILPQALRRMLPPLAS	
30			
•	orf129-1	SLALIANSGAYICEIFRAGIQSIDKGQMEAARSLGLTYPQAMRYVILPQALRRMLPPLAS	
	orf129a.pep	FETMI I PROCEE OUTAWARE AND	
	Offizaa.pep	EFITLLKDSSLLSVIAVAELAYVQNTITGRYSVYEEPLYTVALIYLLMTTFLGWIFLRLE	
35	orf129-1	EFITLLKDSSLLSVIAVAELAYVQNTITGRYSVYEEPLYTVALIYLLMTTFLGWIFLRLE	
50	V11127 1	TITION TO THE DESIGNATION OF THE PROPERTY OF THE DESIGNATION OF THE PROPERTY O	
	orf129a.pep	KRYNPQHRX	
	•		
40	orf129-1	KRYNPQHRX	
40			
	Homology with a p	oredicted ORF from N.gonorrhoeae	
	ODE120 shows 00	00/ identity ever a 00 on eventor with a modisted ODE (ODE) 00	
	ORF 123 SHOWS 98	.9% identity over a 88 aa overlap with a predicted ORF (ORF129ng) fi	[0]
	N.gonorrhoeae:		
	14.gonor nocae.		

45	orf129.pep	IIYEYRWMFLYGALTTLGLTVVAXAGGSVLGLLLALARLIHLEKAGAPMRVLAW	54
73	orf129ng	MDFRFDIIYEYRWMFLYGALTTLGLTVVATAGGSVLGLLLALARLIHLEKAGAPMRVLAW	60
	orf129.pep	ALRKVSLLYVTLFRGTPLFVQIVIWAYVWFPFFV	88
50	orf129ng		120

An ORF129ng nucleotide sequence <SEQ ID 841> was predicted to encode a protein having amino acid sequence <SEQ ID 842>:

- 1 MDFRFDIIYE YRWMFLYGAL TTLGLTVVAT AGGSVLGLLL ALARLIHLEK 51 AGAPMRVLAW ALRKVSLLYV TLFRGTPLFV QIVIWAYVWF PFFVILHTAF 101 LGNAMRQSRR VPDKGRWIAG SLELNCQPRG RKTRGEFPPG ESNLGTEPRN 151 PLSMGQRRFP GCENWYPPQN FIKK\*

55

Further work revealed the following gonococcal sequence <SEQ ID 843>:

1 ATGGATTTTC gtTTTGACAT TATTTAcgaA TACCGCTGGA TGTTTCTTTA

```
CGGCGCACTG Acgacettgg ggctgacggt cgtggcgacg gCGGGCGGTT CGGtattggG TCTGTTGTTG GCGTTGGCGC GCCTGATTCA CTTGGAAAAA
               101
                    GCCGGTGCGC CGATGCGCGT GCTGGCGTGG GCGTTGCGTA AGGTTTCGCT
               151
                    GCTGTACGTT ACCCTGTTCC GGGGTACGCC GCTGTTTGTG CAGATTGTGA
               201
5
               251
                    TTTGGGCGTA TGTGTGGTTT CCGTTTTTCG TCCATCCTTC AGACGGCATT
                    TTGGTCAGCG GCGAGGCGGC AATCGCGCTG CGTCGCGGAT ACGGGCCGCT
               301
                    GATTGCCGGT TCTTTGGCAC TGATCGCCAA CTCGGGGGCG TATATCTGTG
               351
                    AGATTTTCCG CGCGGGCATC CAGTCTATAG ACAAAGGACA GATGGAGGCG
               401
                    GCGTGTTCTT TGGGACTGAC CTATCCGCAG GCGATGCGCT ATGTGATTCT
               451
                    GCCGCAGGCA TTGCGCCGTA TGCTGCCGCC TTTGGCGAGC GAGTTCATCA
10
               501
                    CGCTCTTGAA AGACAGCTCG CTGCTGTCGG TCATTGCTGT GGCGGAGTTG
               551
               601
                    GCGTATGTTC AGAATACGAT TACGGGCCGG TATTCGGTTT ATGAAGAACC
                    GCTTTACACC GCCGCCCTGA TTTATCTGTT GATGACGACT TTCTTAGGCT
               651
                    GGATATTCCT GCGTTTGGAA AAACGTTACA ATCCGCAACA CCGCTGA
     This corresponds to the amino acid sequence <SEQ ID 844; ORF129ng-1>:
15
                 1 MDFRFDIIYE YRWMFLYGAL TTLGLTVVAT AGGSVLGLLL ALARLIHLEK
                51
                    AGAPMRVLAW ALRKVSLLYV TLFRGTPLFV QIVIWAYVWF PFFVHPSDGI
               101 LVSGEAAIAL RRGYGPLIAG SLALIANSGA YICEIFRAGI QSIDKGQMEA
                    ARSLGLTYPQ AMRYVILPQA LRRMLPPLAS EFITLLKDSS LLSVIAVAEL
               151
20
                    AYVONTITGR YSVYEEPLYT VALIYLLMTT FLGWIFLRLE KRYNPQHR*
     ORF129ng-1 and ORF129-1 show 99.2% identity in 248 aa overlap:
          orf129-1.pep MDFRFDIIYEYRWMFLYGALTTLGLTVVATAGGSVLGLLLALARLIHLEKAGAPMRVLAW
                       orf129ng-1
                       MDFRFDIIYEYRWMFLYGALTTLGLTVVATAGGSVLGLLLALARLIHLEKAGAPMRVLAW
25
          orf129-1.pep ALRKVSLLYVTLFRGTPLFVQIVIWAYVWFPFFVHPSDGILVSGEAAIALRRGYGPLIAG
                       orf129ng-1
                       ALRKVSLLYVTLFRGTPLFVQIVIWAYVWFPFFVHPSDGILVSGEAAIALRRGYGPLIAG
30
          orf129-1.pep SLALIANSGAYICEIFRAGIQSIDKGQMEAARSLGLTYPQAMRYVILPQALRRMLPPLAS
                       SLALIANSGAYICEIFRAGIQSIDKGQMEAACSLGLTYPQAMRYVILPQALRRMLPPLAS
           orf129ng-1
           orf129-1.pep EFITLLKDSSLLSVIAVAELAYVQNTITGRYSVYEEPLYTVALIYLLMTTFLGWIFLRLE
35
                       EFITLLKDSSLLSVIAVAELAYVQNTITGRYSVYEEPLYTAALIYLLMTTFLGWIFLRLE
           orf129ng-1
           orf129-1.pep KRYNPQHRX
                       111111111
40
           orf129ng-1
                       KRYNPQHRX
      In addition, ORF129ng-1 is homologous to an ABC transporter from A.fulgidus:
           2650409(AE001090) glutamine ABC transporter, permease protein (glnP)
           [Archaeoglobus fulgidus]Length = 224
            Score = 132 bits (329), Expect = 2e-30
45
            Identities = 86/178 (48%), Positives = 103/178 (57%), Gaps = 18/178 (10%)
           Query: 65 VSLLYVTLFRGTPLFVQIVIWAYVWFPFFVHPSDGILVSGEAAIALRRGYGPLIAGSLAL 124
                      +S YV + RGTPL VQI+I
                                              +F P+ GI + E A
           Sbjct: 58 ISTAYVEVIRGTPLLVQILI-----VYFGLPAIGINLQPEPA----
                                                                     ----GIIAL 99
 50
           Query: 125 IANSGAYICEIFRAGIQSIDKGQMEAACSLGLTYPQAMRYVILPQALRRMLPPLASEFIT 184
                         SGAYI EI RAGI+SI GQMEAA SLG+TY QAMRYVI PQA R +LP L +EFI
           Sbjct: 100 SICSGAYIAEIVRAGIESIPIGQMEAARSLGMTYLQAMRYVIFPQAFRNILPALGNEFIA 159
 55
           Query: 185 LLKDSSLLSVIAVAELAYVQNTITGRYSVYEEPLYTAALIYLLMTTFLGWIFLRLEKR 242
                      LLKDSSLLSVI++ EL V
                                                         AL YL+MT L +
           Sbjct: 160 LLKDSSLLSVISIVELTRVGRQIVNTTFNAWTPFLGVALFYLMMTIPLSRLVAYSQKK 217
      This analysis, including the identification of transmembrane domains in the two proteins, suggests
      that the proteins from N.meningitidis and N.gonorrhoeae, and their epitopes, could be useful
      antigens for vaccines or diagnostics, or for raising antibodies.
 60
```

#### Example 100

The following partial DNA sequence was identified in N. meningitidis <SEQ ID 845>:

```
..CTGAAAGAAT GCCGTCTGAA AGACCCTGTT TTTATTCCAA ATATCGTTTA
                  51
                         TAAGAACATC GCCATTACTT TCCTGCTCTT GCACGCCGCC GCCGAACTTT
5
                         GGCTGCCCGC GCAAACCGCC GGTTTTACCG CGCTCGCCGT CGGCTTCATC
                 101
                151
                         CTGCTCGCCA AGCTGCGTGA GCTTCACCAT CACGAACTCT TACGTAAACA
                         CTACGTCCGC ACTTATTACY TGCTCCAACT CTTTGCCGCC GCAGGCTAGT
TTGTGGACAG GCGCGGCGWA ATTACAAAAC CTGCCCGCYT CCGCGCCCCT
                 201
                 251
                 301
                         GCACCTGATT ACCCTCGGCG GCATGATGGG CGGCGTGATG ATGGTGTGGC
10
                 351
                         TGACCGCCGG ACTGTGGCAC AGCGGCTTTA CCAAACTCGA CTACCCCAAA
                 401
                         CTCTGCCGCA TTGCCGTCCC CATCCTTTTC GCCGCCGCCG TCTCGCGCGC
                         TTTCTTGTTG AACGTGAACC CGTTATTTTT CATTACCGTT CCTGCGATTC
                 451
                 501
                         TGACCGCCGC CGTATTCGTA CTGTATCTTT TCrCGTTTAT ACCGATATTT
                         CGGGCGAATG CGTTTACAGA CGATCCGGAr TAr
                 551
```

15 This corresponds to the amino acid sequence <SEQ ID 846; ORF130>:

```
1 ...LKECRLKDPV FIPNIVYKNI AITFLLLHAA AELWLPAQTA GFTALAVGFI
51 LLAKLRELHH HELLRKHYVR TYYLLQLFAA AGSLWTGAAX LQNLPASAPL
101 HLITLGGMMG GVMMVWLTAG LWHSGFTKLD YPKLCRIAVP ILFAAAVSRA
151 FLXNVNPXFF ITVPAILTAA VFVLYLFXFI PIFRANAFTD DPE*
```

20 Further work revealed the complete nucleotide sequence <SEQ ID 847>:

```
ATGCGGCCGT TTTTCGTCGG CGCGGCGGTG CTTGCCATAC TCGGTGCGCT
                 51
                     GGTGTTTTTC ATCAACCCCG GTGCCATCGT CCTGCACCGC CAAATTTTCT
                     TGGAACTTAT GCTGCCGGCG GCATACGGCG GTTTTTTGAC TGCGGCTTTG
                101
                     TTGGACTGGA CGGGTTTTTC GGGTAACCTG AAACCTGTCG CGACTTTGAT
25
                     GGCGGCATTA TTGCTCGCCG CATCCGCTAT ACTGCCCTTT TCGCCGCAAA CTGCCTCGTT TTTCGTCGCC GCCTATTGGC TGGTGTTGCT GCTGTTCTGC
                201
                251
                     GCCCGGCTGA TTTGGCTAGA CCGAAACACC GACAACTTCG CCCTGCTAAT
                301
                     GTTACTTGCC GCGTTCACTG TTTTTCAGAC GGCATATGCC GTCAGCGGCG
                351
                     ATTTGAACCT GTTGCGCGCG CAAGTGCATC TAAATATGGC GGCGGTGATG
                401
30
                     TTCGTATCCG TGCGCGTCAG TATTCTTTTG GGCGCGGAAG CCCTGAAAGA
                451
                501
                     ATGCCGTCTG AAAGACCCTG TTTTTATTCC AAATATCGTT TATAAAAACA
                     TCGCCATTAC TTTCCTGCTC TTGCACGCCG CCGCCGAACT TTGGCTGCCC
                551
                601
                     GCGCAAACCG CCGGTTTTAC CGCGCTCGCC GTCGGCTTCA TCCTGCTCGC
                     CAAGCTGCGT GAGCTTCACC ATCACGAACT CTTACGTAAA CACTACGTCC
                651
35
                701
                     GCACTTATTA CCTGCTCCAA CTCTTTGCCG CCGCAGGCTA TTTGTGGACA
                     GGCGCGGCGA AATTACAAAA CCTGCCCGCC TCCGCGCCCC TGCACCTGAT
                751
                801
                     TACCCTCGGC GGCATGATGG GCGGCGTGAT GATGGTGTGG CTGACCGCCG
                     GACTGTGGCA CAGCGGCTTT ACCAAACTCG ACTACCCCAA ACTCTGCCGC
                851
                     ATTGCCGTCC CCATCCTTTT CGCCGCCGCC GTCTCGCGCG CTTTCTTGAT
                901
40
                     GAACGTGAAC CCGATATTTT TCATTACCGT TCCTGCGATT CTGACCGCCG
               1001
                     CCGTATTCGT ACTGTATCTT TTCACGTTTA TACCGATATT TCGGGCGAAT
                     GCGTTTACAG ACGATCCGGA ATAA
               1051
```

This corresponds to the amino acid sequence <SEQ ID 848; ORF130-1>:

```
45 S1 MRPFFVGAAV LAILGALVFF INPGAIVLHR QIFLELMLPA AYGGFLTAAL

51 LDWTGFSGNL KPVATLMAAL LLAASAILPF SPQTASFFVA AYWLVLLLFC

101 ARLIWLDRNT DNFALLMLLA AFTVFQTAYA VSGDLNLLRA QVHLNMAAVM

151 FVSVRVSILL GAEALKECRL KDPVFIPNIV YKNIAITFLL LHAAAELWLP

201 AQTAGFTALA VGFILLAKLR ELHHHELLRK HYVRTYYLLQ LFAAAGYLWT

251 GAAKLQNLPA SAPLHLITLG GMMGGVMMVW LTAGLWHSGF TKLDYPKLCR

50 301 IAVPILFAAA VSRAFLMNVN PIFFITVPAI LTAAVFVLYL FTFIPIFRAN

351 AFTDDPE*
```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from N. meningitidis (strain A)

ORF130 shows 94.3% identity over a 193aa overlap with an ORF (ORF130a) from strain A of N.

55 meningitidis:

				10	20	30
	orf130.pep				FIPNIVYKNIAI	
_	orf130a LN	LLRAQVHLNMAAVMFV		ALKECRLKDPVI	PIPNVVYKNIAI	TFLLLHAA
5		140 150	160	170	180	190
			50 6		80	90
		LWLPAQTAGFTALAVG				
10	orf130a AE	LWLPAQTAGFTSLAVG	FILLAKLRELH	HHELLRKHYVR'	ryyllolfaaag	YLWTGAAK
		200 210	220	230	240	250
			10 12		140	150
15		NLPASAPLHLITLGGM 				
10		NLPASAPLHLITLGGM	MGSVMMVWLTA	GLWHSGFTKLD'	YPKLCRIAVPIL	FAAAVSRA
		260 270	280	290	300	310
20	orf130.pep FI	160 1 LXNVNPXFFITVPAILT	70 18		DDEY	
20						
	orf130a VI	LMNVNPIFFITVPAILT 320 330	AAVFVLYLLTF 340	VPIFRANAFTD 350	DPEX	
	m 1.4. 1				· :	
	The complete length C	JKF 130a nucleotia	e sequence <	SEQ ID 849	> 1S:	
25		GCCGT TTTTCGTCGG				'
		PTTTTC ATCAACCCCG ACTTAT GCTGCCGGCG				•
		ACTGGA CGGGTTTTTC				
20		SCATTA TTGCTCGCCG				
30		CTCGTT TTTCGTCGCC GGCTGA TTTGGCTAGA				
		CTTGCC GCGTTCACTG				
		GAACCT GTTGCGCGCG				
35		PATCCG TGCGCGTCAG CGTCTG AAAGACCCAG				
33		CATTAC CTTCCTGCTC				
		AAACCG CCGGTTTTAC				
		CTGCGT GAGCTTCACC ITATTA CCTGCTCCAA				
40	751 GGCGG	CGGCGA AATTACAAAA	CCTGCCGCC	TCCGCGCCCC	TGCACCTGAT	
		CTCGGT GGCATGATGG				
		GTGGCA CAGCGGCTTT CCGTCC CCATCCTNTT				
		GTAAAC CCGATATTCT				
45		GTTCGT GCTTTACCTG TTACAG ACGATCCGGA		TACCGATCTT	TCGGGCGAAC	
				-ano m oco		
	This encodes a proteir	n naving amino aci	a sequence <	(2EQ ID 850	<i>)&gt;</i> :	
		FVGAAV LAILGALVFE				
60	51 LDWT	GFSGNL KPVATLMAAI	LLAASAILP	SPQTASFFVA	AYWLVLLLFC	
50		WLDRNT DNFALLMLLA RVSILL GAEALKECRI				
	201 AQTA	GFTSLA VGFILLAKLE	RELHHHELLRE	K HYVRTYYLLC	LFAAAGYLWT	
	251 GAAK	LONLPA SAPLHLITLO	GMMGSVMMVV	V LTAGLWHSGE	TKLDYPKLCR	
55	301 <u>IAVP</u> 351 AFTD	<u>ILFAAA VSRAVLM</u> NVN DPE*	PIFFITVPA	L LTAAVEVLYL	LIEVPIERAN	
	ORF130a and ORF13	60-1 show 98.3% id	lentity in 35	7 aa overlap:		
			•	_		
	orf130a.pep	MRPFFVGAAVLAILO				
<b>C</b> 0	orf130-1	MRPFFVGAAVLAIL				
60	orf130a.pep	KPVATLMAALLLAA:	SAILPFSPQTA	SFFVAAYWLVLI	LFCARLIWLDR	NTONFALLMLLA
		111111111111111	HILLIA	111111111111		11111111111
	orf130-1	KPVATLMAALLLAA:	SALLPFSPQTA:	SEEVAAIWLVLI	ine CAKPIMPDK	NIUNIALLMLLA
65	orf130a.pep	AFTVFQTAYAVSGD	LNLLRAQVHLN	MAAVMFVSVRVS	SILLGAEALKEC	RLKDPVFIPNVV

60

	orf130-1	
5	orf130a.pep	YKNIAITFLLLHAAAELWLPAQTAGFTSLAVGFILLAKLRELHHHELLRKHYVRTYYLLQ
5	orf130-1	
	orf130a.pep	LFAAAGYLWTGAAKLQNLPASAPLHLITLGGMMGSVMMVWLTAGLWHSGFTKLDYPKLCR
10	orf130-1	
	orf130a.pep	IAVPILFAAAVSRAVLMNVNPIFFITVPAILTAAVFVLYLLTFVPIFRANAFTDDPE
15	orf130-1	
13	Homology with a p	redicted ORF from N.gonorrhoeae
	ORF130 shows 91.	7% identity over a 193 aa overlap with a predicted ORF (ORF130ng) from
	N.gonorrhoeae:	
20	orf130.pep	LKECRLKDPVFIPNIVYKNIAITFLLLHAA 30
20	orf130ng	:::
	orf130.pep	AELWLPAQTAGFTALAVGFILLAKLRELHHHELLRKHYVRTYYLLQLFAAAGSLWTGAAX 90
25	orf130ng	AELWLPAQTAGFTALAVGFILLAKLRELHHHELLRKHYVRTYYLLQLFAAAGYLWTGAAK 261
	orf130.pep	LQNLPASAPLHLITLGGMMGGVMMVWLTAGLWHSGFTKLDYPKLCRIAVPILFAAAVSRA 150
30	orf130ng	LQNLPASAPLHLITLGGMTGGVMMVWLTAGLWHSGFTKLDYPKLCRIAVSILFASAVSRA 321
	orf130.pep	FLXNVNPXFFITVPAILTAAVFVLYLFXFIPIFRANAFTDDPE 193
	orf130ng	VLMNVNPIFFITVPEILTAAVFMLYLLTFVPIFRANAFTDDPE 364
	An ORF130ng nucl	eotide sequence <seq 851="" id=""> was predicted to encode a protein having amino</seq>
35	acid sequence <se< th=""><th>Q ID 852&gt;:</th></se<>	Q ID 852>:
		KFFTHPMR PFFVGAAVLA ILGALVFFHQ PRRYHPAPPN FLGTYAAGCI FFDYRFVG PDGFFRQPET CRYFDGGVVA CCGCFIAVFT ATCRIFRRRL
	101 LA	GVAAVLRL ADLARROHRT LRSVDVTAAF TVFOTAYAVS GDLNLLRAOV
40	151 H <u>I</u> 201 AA	NMAAVMFV SVRVSVLLGT ETLKECRLKD PVFIPNVIYK NIAITLLLHA ELWLPAQT AGFTALAVGF ILLAKLRELH HHELLRKHYV RTYYLLQLFA
	251 AF	GYLWTGAA KLONLPASAP LHLITLGGMT GGVMMVWLTA GLWHSGFTKL PKLCRIAV SILFASAVSR AVLMNVNPIF FITVPEILTA AVFMLYLLTF
	351 <u>V</u> E	IFRANAFT DDPE*
	Further work revea	led the following gonococcal DNA sequence <seq 853="" id="">:</seq>
45	1 AT	GCGCCCGT TTTTCGTCGG TGCGGCAGTA CTTGCCATAC TCGGTGCGTT
	51 GG 101 TG	TGTTTTTT ATCAACCCCG GCGCTATCAT CCTGCACCGC CAAATTTTCT GAACTTAT GCTGCCGGCT GCATACGGCG GTTTTTTGAC TACCGCTTTG
	151 TT	GGACCGGA CGGGTTTTTC AGGCAACCTG AAACCTGCCG CTACTTTGAT
50		CGGTGTTG TTGCTTGTTG CGGCTGTTTT ATTGCCGTTT TTACCGCAAC CGCCGCATT TTTCGTCGCC GCCTATTGGC TGGTGTTGCT GCTGTTCTGC
	301 GC	CTGGCTGA TTTGGCTCGA CCGCAACACC GACAACTTCG CTCTGTTGAT
	351 G1	TACTTGCC GCATTTACCG TTTTTCAGAC GGCCTATGCC GTCAGCGGCG
	401 AT 451 TT	TTGAACTT ACTGCGCGC CAAGTGCATT TGAATATGGC GGCGGTCATG CGTATCCG TCCGCGTCAG CGTCCTTTTG GGCACGGAAA CCCTGAAAGA
55	501 AT	GCCGTCTG AAAGACCCCG TATTCATCCC CAACGTTATC TATAAAAACA
	551 TC	GCCATCAC CCTGCTGCTG CACGCCGCCG CCGAACTTTG GCTGCCCGCG

551 TCGCCATCAC CCTGCTGCTG CACGCCGCCG CCGAACTTTG GCTGCCCGCG
601 CAAACCGCCG GTTTTACTGC GCTTGCCGTC GGCTTCATCC TGCTCGCCAA
651 GCTGCGCGAA CTGCACCATC ACGAACTCTT ACGCAAACAC TACGTCCGCA 701 CTTATTACCT GCTCCAGCTC TTTGCCGCCG CAGGTTATCT GTGGACAGGC 751 GCGGCGAAAC TGCAAAACCT GCCCGCCTCC GCGCCCCTGC ACCTGATTAC 801 CCTCGGCGGC ATGACGGGTG GCGTGATGAT GGTGTGGCTG ACTGCCGGAC 851 TGTGGCACAG CGGCTTTACC AAACTCGACT ACCCGAAACT CTGCCGCATC

	GCCGTCTCCA				
951	CGTGAATCCG	ATATTCTTCA	TCACCGTTCC	CGAGATTCTG	ACCGCCGCCG
1001	TGTTCATGCT	TTACCTGCTG	ACGTTCGTAC	CGATTTTTCG	AGCGAACGCG
1051	TTTACAGACG	ATCCGGAATA	A		

## 5 This corresponds to the amino acid sequence <SEQ ID 854; ORF130ng-1>:

```
1 MRPFFVGAAV LAILGALVFF INPGAIILHR QIFLEMLPA AYGGFLTTAL
51 LDRTGFSGNL KPAATLMAVL LLVAAVLLPF LPQLAAFFVA AYWLVLLLFC
101 AWLIWLDRNT DNFALLMLLA AFTVFQTAYA VSGDLNLLRA QVHLNMAAVM
151 FVSVRVSVLL GTETLKECRL KDPVFIPNVI YKNIAITLLL HAAAELWLPA
10 201 QTAGFTALAV GFILLAKLRE LHHHELLRKH YVRTYYLLQL FAAAGYLWTG
251 AAKLQNLPAS APLHLITLGG MTGGVMVVUL TAGLWHSGFT KLDYPKLCRI
301 AVSILFASAV SRAVLMNVNP IFFITVPEIL TAAVFMLYLL TFVPIFRANA
351 FTDDPE*
```

#### ORF130ng-1 and ORF130-1 show 92.4% identity in 357 aa overlap:

15	orf130-1.pep orf130ng-1	MRPFFVGAAVLAILGALVFFINPGAIVLHRQIFLELMLPAAYGGFLTAALLDWTGFSGNL 
20	orf130-1.pep orf130ng-1	KPVATLMAALLLAASAILPFSPQTASFFVAAYWLVLLLFCARLIWLDRNTDNFALLMLLA     :
25	orf130-1.pep orf130ng-1	AFTVFQTAYAVSGDLNLLRAQVHLNMAAVMFVSVRVSILLGAEALKECRLKDPVFIPNIV 
20	orf130-1.pep orf130ng-1	YKNIAITFLLLHAAAELWLPAQTAGFTALAVGFILLAKLRELHHHELLRKHYVRTYYLLQ 
30	orf130-1.pep orf130ng-1	LFAAAGYLWTGAAKLQNLPASAPLHLITLGGMMGGVMMVWLTAGLWHSGFTKLDYPKLCR 
35	orf130-1.pep orf130ng-1	IAVPILFAAAVSRAFLMNVNPIFFITVPAILTAAVFVLYLFTFIPIFRANAFTDDPEX         :

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### 40 Example 101

55

The following partial DNA sequence was identified in N.meningitidis <SEQ ID 855>:

```
ATGGAAATTC GGGCAATAAA ATATACGCCA ATGGCTGCGT TGCTTGCATT

51 TACGGTTGCA GGCTGCCGGC TGGCGGGGTG GTATGAGTGT TCGTCCCTCA

101 CCGGCTGGTG TAAGCCGAGA AAACCGGCTG CCATCGATTT TTGGGATATT

151 GGCGGCGAGA GTCCGCCGTC TTTAGGGGAC TACGACATAC CGCTTTCAGA

201 CGGCAATAGT TCCGTCAGGG CAAACGAATA TGAATCCGCA CAACAATCTT

251 ACTTTTACAG GAAAATAGGG AAGTTTGAAG C.TGCGGGCT GGATTGGCGT

301 ACGCGTGACG GCAAACCTTT GATTGAGACG TTCAAACAGG GAGGATTTGA

351 CTGCTTGGAA AAG..
```

#### 50 This corresponds to the amino acid sequence <SEQ ID 856; ORF131>:

- 1 MEIRAIKYTA MAALLAFTVA GCRLAGWYEC SSLTGWCKPR KPAAIDFWDI
- 51 GGESPPSLGD YEIPLSDGNS SVRANEYESA QQSYFYRKIG KFEXCGLDWR
- 101 TRDGKPLIET FKQGGFDCLE K..

Further work revealed the complete nucleotide sequence <SEQ ID 857>:

1 ATGGAAATTC GGGCAATAAA ATATACGGCA ATGGCTGCGT TGCTTGCATT 51 TACGGTTGCA GGCTGCCGGC TGGCGGGGTG GTATGAGTGT TCGTCCCTCA

55

orf131a.pep

	-466-	
5	CCGGCTGGTG TAAGCCGAGA AAACCGGCTG CCATCGATTT TTGGGATATT  151 GGCGCGAGA GTCCGCCGTC TTTAGGGGAC TACGAGATAC CGCTTTCAGA  201 CGGCAATCGT TCCGTCAGGG CAAACGAATA TGAATCCGCA CAACAATCTT  251 ACTTTACAG GAAAATAGGG AAGTTTGAAG CCTGCGGGCT GGATTGGCGT  301 ACGCGTGACG GCAAACCTTT GATTGAGACG TTCAAACAGG GAGGATTTGA  351 CTGCTTGGAA AAGCAGGGGT TGCGGCGCAA CGGTCTGTCC GAGCGCGTCC  401 GATGGTAA	
	This corresponds to the amino acid sequence <seq 858;="" id="" orf131-1="">:</seq>	
10	1 MEIRAIKYTA MAALLAFTVA GCRLAGWYEC SSLTGWCKPR KPAAIDFWDI 51 GGESPPSLGD YEIPLSDGNR SVRANEYESA QQSYFYRKIG KFEACGLDWR 101 TRDGKPLIET FKQGGFDCLE KQGLRRNGLS ERVRW*	
	Computer analysis of this amino acid sequence gave the following results:	
	Homology with a predicted ORF from N.meningitidis (strain A)	
	ORF131 shows 95.0% identity over a 121aa overlap with an ORF (ORF131a) from strain A of N.	
15	meningitidis:	
	10 20 30 40 50 60  orf131.pep MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSLTGWCKPRKPAAIDFWDIGGESPPSLGD	
20	10 20 30 40 50 60	
	70 80 90 100 110 120 orf131.pep YEIPLSDGNSSVRANEYESAQQSYFYRKIGKFEXCGLDWRTRDGKPLIETFKQGGFDCLE	
25	orf131a YEIPLSDGNRSVRANEYESAQQSYFYRKIGKFEACGLDWRTRDGKPLIETFKQEGFDCLK 70 80 90 100 110 120	
•		
30	orf131.pep K     orf131a KQGLRRNGLSERVRWX	
	The complete length ORF131a nucleotide sequence <seq 859="" id=""> is:</seq>	
35	1 ATGGAAATTC GGGCAATAAA ATATACGGCA ATGGCTGCGT TGCTTGCATT 51 TACGGTTGCA GGCTGCCGGT TGGCAGGTTG GTATGAGTGT TCGTCCCTGT 101 CCGGCTGGTG TAAGCCGAGA AAACCTGCCG CCATCGATTT TTGGGATATT 151 GGCGGCGAGA GTCCTCCGTC TTTAGAGGAC TACGAGATAC CGCTTTCAGA 201 CGGCAATCGT TCCGTCAGGG CAAACGAATA TGAATCCGCA CAACAATCTT	
40	251 ACTTTTACAG GAAAATAGGG AAGTTTGAAG CCTGCGGGTT GGATTGGCGT 301 ACGCGTGACG GCAAACCTTT GATTGAGACG TTCAAACAGG AAGGTTTTGA 351 TTGTTTGAAA AAGCAGGGGT TGCGGCGCAA CGGTCTGTCC GAGCGCGTCC 401 GATGGTAA	٠
	This encodes a protein having amino acid sequence <seq 860="" id="">:</seq>	
45	1 MEIRAIKYTA MAALLAFTVA GCRLAGWYEC SSLSGWCKPR KPAAIDFWDI 51 GGESPPSLED YEIPLSDGNR SVRANEYESA QQSYFYRKIG KFEACGLDWR 101 TRDGKPLIET FKQEGFDCLK KQGLRRNGLS ERVRW*	
	ORF131a and ORF131-1 show 97.0% identity in 135 aa overlap:	
	orf131a.pep MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSLSGWCKPRKPAAIDFWDIGGESPPSLED	
50	orf131-1 MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSLTGWCKPRKPAAIDFWDIGGESPPSLGD	
	orf131a.pep YEIPLSDGNRSVRANEYESAQQSYFYRKIGKFEACGLDWRTRDGKPLIETFKQEGFDCLK	
55	orf131-1 YEIPLSDGNRSVRANEYESAQQSYFYRKIGKFEACGLDWRTRDGKPLIETFKQGGFDCLE	

KQGLRRNGLSERVRWX

-467-

```
orf131-1 KQGLRRNGLSERVRWX
```

#### Homology with a predicted ORF from N.gonorrhoeae

ORF131 shows 89.3% identity over 121 aa overlap with a predicted ORF (ORF131ng) from

5 N.gonorrhoeae:

50

```
60
                  MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSLTGWCKPRKPAAIDFWDIGGESPPSLGD
        orf131.pep
                  MEIRVIKYTATAALFAFTVAGCRLAGWYECLSLSGWCKPRKPAAIDFWDIGGESPLSLED
        orf131ng
10
                  YEIPLSDGNSSVRANEYESAQQSYFYRKIGKFEXCGLDWRTRDGKPLIETFKQGGFDCLE
        orf131.pep
                  YEIPLSDGNRSVRANEYESAQKSYFYRKIGKFEACGLDWRTRDGKPLVERFKQEGFDCLE
        orf131ng
                              121
        orf131.pep
                  K
15
                  KQGLRRNGLSERVRW
                              134
        orfl31ng
```

A complete length ORF131ng nucleotide sequence <SEQ ID 861> was predicted to encode a protein having amino acid sequence <SEQ ID 862>:

```
1 MEIRVIKYTA TAALFAFTVA GCRLAGWYEC LSLSGWCKPR KPAAIDFWDI
20 51 GGESPLSLED YEIPLSDGNR SVRANEYESA QKSYFYRKIG KFEACGLDWR
101 TRDGKPLVER FKQEGFDCLE KQGLRRNGLS ERVRW*
```

Further work revealed the following gonococcal DNA sequence <SEQ ID 863>:

```
ATGGAAATTC GGGTAATAAA ATATACGGCA ACGGCTGCGT TGTTTGCATT
                   51
                       TACGGTTGCA GGCTGCCGGC TGGCGGGGTG GTATGAGTGT TCGTCCTTGT
25
                       CCGGCTGGTG TAAGCCGAGA AAACCTGCCG CCATCGATTT TTGGGATATT
                  101
                       GGCGGCGAGA GtccgctGTC TTTAGAGGAC TACGAGATAC CGCTTTCAGA
CGGCAATCGT TCCGTCAGGG CAAACGAATA TGAATCCGCG CAAAAATCTT
                  151
                  201
                       ACTTTTATAG GAAAATAGGG AAGTTTGAAG CCTGCGGGTT GGATTGGCGT
                  251
                       ACGCGTGACG GCAAACCTTT GGTTGAGAGG TTCAAACAGG AAGGTTTCGA
                  301
30
                  351
                       CTGTTTGGAA AAGCAGGGGT TGCGGCGCAA CGGCCTGTCC GAGCGCGTCC
                  401
                       GATGGTAA
```

This corresponds to the amino acid sequence <SEQ ID 864; ORF131ng-1>:

```
1 MEIRVIKYTA TAALFAFTVA GCRLAGWYEC SSLSGWCKPR KPAAIDFWDI
51 GGESPLSLED YEIPLSDGNR SVRANEYESA QKSYFYRKIG KFEACGLDWR
35 101 TRDGKPLVER FKQEGFDCLE KQGLRRNGLS ERVRW*
```

ORF131ng-1 and ORF131-1 show 92.6% identity in 135 aa overlap:

Based on the presence of a predicted prokaryotic membrane lipoprotein lipid attachment site, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 102

The following partial DNA sequence was identified in N.meningitidis <SEQ ID 865>

```
ATGAAACACA TCCATATTAT CGGTATCGGC GGCACGTTTA TGGGCGGGCT
                       TGCCGCCATT GCCAAAGAAG CGGGGTTTGA AGTCAGCGGT TGCGACGCGA
 5
                       AGATGTATCC GCCGATGAGC ACCCAGCTCG AAGCCTTGGG TATAGACGTG
                 101
                      TATGAAGGCT TCGATGCCGC TCAGTTGGAC GAATTTAAAG CCGACGTTTA
                      CGTTATCGGC AATGTCGCCA AGCGCGGGAT GGATGTGGTT GAAGCGATTT
                      TGAACCTCGG CCTGCCTTAT ATTTCCGGCC CGCAATGGCT GTCGGAAAAC GTGCTGCACC ATCATTGGGT ACTCGGTGTG GCGGGGACGC ACGGCAAAAC
                 301
                 351 GACCACCGCC TCCATGCTCG CATGGGTCTT GGAATATGCC GGCCTCGCGC 401 CGGGCTTCCT TATLGGCGGC GTACC.GGAA AATLLCGGCG TTTCCGCCCG
10
                 451
                      CCTGCCGCAA ACGCCGCGCC AAGACCCGAA CAGCCAATCG CCGTTTTTcG
                 501 TCATCGAAGC CGACGAATAC GACACCGCCT TTLTCGACAA ACGTTCTAAA
551 TLCGTGCATT ACCGTCCGCG TACCGCCGTG TTGAACAATC TGGAATTCGA
15
                 601
                      CCACGCCGAC ATCTTTGCCG ACTTGGGCGC GATACAGACC CAGTTCCACT
                      ACCTCGTGCG TACCGTGCCG TCTGAAGGCT TAATCGTCTG CAACGGACGG CAGCAAAGCC TGCAAGATAC TTTGGACAAA GGCTGCTGGA CGCCGGTGGA
                 651
                 701
                 751 AAAATTCGGC ACGGAACACG GCTGGCA..
      This corresponds to the amino acid sequence <SEQ ID 866; ORF132>:
20
                       MKHIHIIGIG GTFMGGLAAI AKEAGFEVSG CDAKMYPPMS TQLEALGIDV
                      YEGFDAAQLD EFKADVYVIG NVAKRGMDVV EAILNLGLPY ISGPQWLSEN
                       VLHHHWVLGV AGTHGKTTTA SMLAWVLEYA GLAPGFLIGG VXGKFRRFRP
                      PAANAAPRPE QPIAVFRHRS RRIRHRLFRQ TFXIRALPSA YRRVEQSGIR
                       PRRHLCRLGR DTDPVPLPRA YRAVXRLNRL QRTAAKPARY FGORLLDAGG
25
                 251 KIRHGTRLA..
      Further work revealed the complete nucleotide sequence <SEQ ID 867>:
                      ATGAAACACA TCCATATTAT CGGTATCGGC GGCACGTTTA TGGGCGGGCT
TGCCGCCATT GCCAAAGAAG CGGGGTTTGA AGTCAGCGGT TGCGACGCGA
                       AGATGTATCC GCCGATGAGC ACCCAGCTCG AAGCCTTGGG TATAGACGTG
30
                      TATGAAGGCT TCGATGCCGC TCAGTTGGAC GAATTTAAAG CCGACGTTTA
CGTTATCGGC AATGTCGCCA AGCGCGGGAT GGATGTGGTT GAAGCGATTT
                       TGAACCTCGG CCTGCCTTAT ATTTCCGGCC CGCAATGGCT GTCGGAAAAC
                 251
                       GTGCTGCACC ATCATTGGGT ACTCGGTGTG GCGGGGACGC ACGGCAAAAC
                  351
                       GACCACCGCC TCCATGCTCG CATGGGTCTT GGAATATGCC GGCCTCGCGC
35
                       CGGGCTTCCT TATTGGCGGC GTACCGGAAA ATTTCGGCGT TTCCGCCCGC
                       CTGCCGCAAA CGCCGCCCA AGACCCGAAC AGCCAATCGC CGTTTTTCGT
                  451
                       CATCGAAGCC GACGAATACG ACACCGCCTT TTTCGACAAA CGTTCTAAAT
                  551
                       TCGTGCATTA CCGTCCGCGT ACCGCCGTGT TGAACAATCT GGAATTCGAC
                  601
                       CACGCCGACA TCTTTGCCGA CTTGGGCGCG ATACAGACCC AGTTCCACTA
40
                       CCTCGTGCGT ACCGTGCCGT CTGAAGGCTT AATCGTCTGC AACGGACGGC
                  651
                       AGCAAAGCCT GCAAGATACT TTGGACAAAG GCTGCTGGAC GCCGGTGGAA
                       AAATTCGGCA CGGAACACGG CTGGCAGGCC GGCGAAGCCA ATGCCGACGG
                  751
                  801
                       CTCGTTCGAC GTGTTGCTCG ACGGCAAAAC CGCCGGACGC GTCAAATGGG
                  851 ATTTGATGGG CAGGCACAAC CGCATGAACG CGCTCGCCGT CATTGCCGCC
45
                       GCGCGTCATG TCGGTGTCGA TATTCAGACC GCCTGCGAAG CCTTGGGCGC
                  951
                       GTTTAAAAAC GTCAAACGCC GGATGGAAAT CAAAGGCACG GCAAACGGCA
                 1001
                       TCACCGTTTA CGACGACTTC GCCCACCACC CGACCGCCAT CGAAACCACG
                      ATTCAAGGTT TGCGCCAACG CGTCGGCGGC GCGCGCATCC TCGCCGTCCT
                 1051
                       CGAACCGCGT TCCAACACGA TGAAGCTGGG CACGATGAAG TCCGCCCTGC
                 1101
50
                       CTGTAAGCCT CAAAGAAGCC GACCAAGTGT TCTGCTACGC CGGCGGCGTG
                 1151
                 1201
                       GACTGGGACG TCGCCGAAGC CCTCGCGCCT TTGGGCGGCA GGCTGAACGT
                 1251
                       CGGCAAAGAC TTCGATGCCT TCGTTGCCGA AATCGTGAAA AACGCCGAAG
                       TAGGCGACCA TATTTTGGTG ATGAGCAACG GCGGTTTCGG CGGAATACAC
                 1301
                       GGAAAGCTGC TGGAAGCTTT GAGATAG
                 1351
 55
       This corresponds to the amino acid sequence <SEQ ID 868; ORF132-1>:
                       MKHIHIIGIG GTFMGGLAAI AKEAGFEVSG CDAKMYPPMS TQLEALGIDV
                       YEGFDAAQLD EFKADVYVIG NVAKRGMDVV EAILNLGLPY ISGPOWLSEN
                       VLHHHWVLGV AGTHGKTTTA SMLAWVLEYA GLAPGFLIGG VPENFGVSAR
                       LPQTPRQDPN SQSPFFVIEA DEYDTAFFDK RSKFVHYRPR TAVLNNLEFD
                  151
 60
                  201
                       HADIFADLGA IQTQFHYLVR TVPSEGLIVC NGRQQSLQDT LDKGCWTPVE
                       KFGTEHGWQA GEANADGSFD VLLDGKTAGR VKWDLMGRHN RMNALAVIAA
```

ARHVGVDIQT ACEALGAFKN VKRRMEIKGT ANGITVYDDF AHHPTAIETT

351 IQGLRQRVGG ARILAVLEPR SNTMKLGTMK SALPVSLKEA DQVFCYAGGV

401 DWDVAEALAP LGGRLNVGKD FDAFVAEIVK NAEVGDHILV MSNGGFGGIH

451 GKLLEALR\*

Computer analysis of this amino acid sequence gave the following results:

5 Homology with the hypothetical o457 protein of E.coli (accession number U14003)

ORF132 and o457 show 58% aa identity in 140 aa overlap:

```
IHIIGIGGTFMGGLAAIAKEAGFEVSGCDAKMYPPMSTQLEALGIDVYEGFDAAQLDEFK 63
IHI+GI GTFMGGLA +A++ G EV+G DA +YPPMST LE GI++ +G+DA+QL+ +
                          IHILGICGTFMGGLAMLARQLGHEVTGSDANVYPPMSTLLEKQGIELIQGYDASQLEP-Q 61
            0457:
                     3
10
            Orf132: 64 ADVYVIGNVAKRGMDVVEAILNLGLPYISGPQWLSENVLHHHWVLGVAGTHGKTTTASML 123
                                                                         WVL VAGTHGKTTTA M
                                                    +PY+SGPQWL + VL
                                     RG
                                           VEA+L
                          PDLVIIGNAMTRGNPCVEAVLEKNIPYMSGPQWLHDFVLRDRWVLAVAGTHGKTTTAGMA 121
            0457:
            Orf132: 124 AWVLEYAGLAPGFLIGGVXG 143
15
                           W+LE G PGF+IGGV G
                      122 TWILEQCGYKPGFVIGGVPG 141
            0457:
```

## Homology with a predicted ORF from N. meningitidis (strain A)

ORF132 shows 74.6% identity over a 189aa overlap with an ORF (ORF132a) from strain A of N. meningitidis:

			10	20	30	40	50	60
	orf132.pep	MKHIHII	GIGGTFMGG	LAAIAKEAG	FEVSGCDAKM	YPPMSTQLEA	LGIDVYEGF	'DAAQLD
		1111111		:11111111	11 1111111		111 11111	1:1111
25	orf132a	MKHIHII	GIGGTEMGG	IAAIAKEAG	FEXSGCDAKM	YPPMSTQLEA	LGIGVYEGF	TAQLD
23	<b>V</b>		10	20	30	40	50	60
			70	80	90	100	110	120
	orf132.pep	EFKADV	<b>CVIGNVAKRO</b>	MDVVEAILN	LGLPYISGPC	WLSENVLHHH	WVLGVAGTH	
30		111111	шнин			11:11 1111	1 1111 11	
	orf132a	EFKADV	YVIGNVAKRO	MDVVEAILN	RGLPYISGPÇ	WLAENXLHHH		IGKTTTA
			70	80	90	100	110	120
			130	140	150	160		
35	orf132.pep	SMLAWV	LEYAGLAPGI	FLIGGVXGKI	RRFRPP	<b>ANAAPRPEQ</b> P	·I	AVFR
		111111	11111111					1 1
	orf132a	SMLAWV:	LEYAGLAPGI	FXIGGVPEN	SVSARL-PQ	PRQDPNSQSE	FFVIEADE	<b>YDTAFFD</b>
			130	140	150	160	170	
40		170	180	190	200	210	220	
	orf132.pep	HRSRRI	RHRLFRQTF	KIRALPSAYI	RRVEQSGIRPI	RRHLCRLGRDI	DPVPLPRA	YRAVXRL
		:11: :	::					
	orf132a	KRSKFV	HYRPRTAVL	NNLEFDHAD:	[FADLGAIOT(	<b>OFHHLVRTVPS</b>	SEGLIVCNG	RQQSLQD
	OFILISZA							
	OF1132a	180	190	200	210	220	230	

The complete length ORF132a nucleotide sequence <SEQ ID 869> is:

101 AGATGTATCC GCCGATGAGC ACCCAGCTCG AAGCCTTGGG CATAGGCGTG 151 TATGAAGGCT TCGACACCGC GCAGTTGGAC GAATTTAAAG CCGACGTTTA 201 CGTTATCGGC AATGTCGCCA AGCGCGGGAT GGATGTGGTT GAAGCGATTT 251 TGAACCGTGG GCTGCCTTAT ATTTCCGGCC CGCAATGGCT GGCTGAAAAC 301 NTGCTGCACC ATCATTGGNN ACTCGGCGTG GCGGNGACGC ACGGCAAAAC 351 GACCACCGCG TCTATGCTCG CGTGGGTTTT GGAATATGCC GGACTCGCAC 401 CGGGCTTCNT TATCGGCGGC GTACCGGAAA ACTTCAGCGT TTCCGCCCGC 401 CTGCCGCAAA CGCCGCGCCA AGACCCGAAC AGCCAATCGC CGTTTTTCGT 551 CATTGAAGCC GACGAATACG ACACCGCGTT TTTCGACAAA CGCTCCAAAT 551 TCGTGCATTA CCGTCCGCGT ACCGCCGTGT TGAACAATCT GGAATTCGACCA 601 CACGCCGACA TCTTCGCCGA TTTGGGCGCG ATACAGACCC AGTTCCACCA 651 CCTCGTGCGT ACCGTGCCGT CTGAAGGCCT CATCGTCTGC AACGGCACGCGCCCCCCCCCC							
101 AGATGTATCC GCCGATGAGC ACCCAGCTCG AAGCCTTGGG CATAGGCGTG 151 TATGAAGGCT TCGACACCGC GCAGTTGGAC GAATTTAAAG CCGACGTTTA 201 CGTTATCGGC AATGTCGCCA AGCGCGGGAT GGATGTGGTT GAAGCGATTT 251 TGAACCGTGG GCTGCCTTAT ATTTCCGGCC CGCAATGGCT GGCTGAAAAC 301 NTGCTGCACC ATCATTGGNN ACTCGGCGTG GCGGNGACGC ACGGCAAAAC 351 GACCACCGCG TCTATGCTCG CGTGGGTTTT GGAATATGCC GGACTCGCAC 401 CGGGCTTCNT TATCGGCGGC GTACCGGAAA ACTTCAGCGT TTCCGCCCGC 401 CTGCCGCAAA CGCCGCGCCA AGACCCGAAC AGCCAATCGC CGTTTTTCGT 551 CATTGAAGCC GACGAATACG ACACCGCGTT TTTCGACAAA CGCTCCAAAT 551 TCGTGCATTA CCGTCCGCGT ACCGCCGTGT TGAACAATCT GGAATTCGACCA 601 CACGCCGACA TCTTCGCCGA TTTGGGCGCG ATACAGACCC AGTTCCACCA 651 CCTCGTGCGT ACCGTGCCGT CTGAAGGCCT CATCGTCTGC AACGGCACGCGCCCCCCCCCC		1	ATGAAACACA	TCCACATTAT	CGGTATCGGC	GGCACGTTTA	TGGGTGGGAT
50  151 TATGAAGGCT TCGACACCGC GCAGTTGGAC GAATTTAAAG CCGACGTTTA 201 CGTTATCGGC AATGTCGCCA AGGCGGGAT GGATGTGGTT GAAGCGATTT 251 TGAACCGTGG GCTGCCTTAT ATTTCCGGCC CGCAATGGCT GGCTGAAAAC 301 NTGCTGCACC ATCATTGGNN ACTCGGCGTG GCGGNGACGC ACGGCAAAAC 351 GACCACCGCG TCTATGCTCG CGTGGGTTTT GGAATATGCC GGACTCGCAC 401 CGGGCTTCNT TATCGGCGGC GTACCGGAAA ACTTCAGCGT TTCCGCCCGC 55 451 CTGCCGCAAA CGCCGCGCCA AGACCCGAAC AGCCAATCGC CGTTTTTCGT 501 CATTGAAGCC GACGAATACG ACACCGCGTT TTTCGACAAA CGCTCCAAAT 551 TCGTGCATTA CCGTCCGCT ACCGCCGTGT TGAACAATCT GGAATTCGACA 601 CACGCCGACA TCTTCGCCGA TTTGGGCGCG ATACAGACCC AGTTCCACCA 651 CCTCGTGCGT ACCGTGCCGT CTGAAGGCCT CATCGTCTGC AACGGCACGCG		51	TGCCGCCATT	GCCAAAGAAG	CAGGGTTTGA	ANTCAGCGGT	TGCGATGCGA
50 201 CGTTATCGGC AATGTCGCCA AGCGCGGGAT GGATGTGGTT GAAGCGATTT 251 TGAACCGTGG GCTGCCTTAT ATTTCCGGCC CGCAATGGCT GGCTGAAAAC 301 NTGCTGCACC ATCATTGGNN ACTCGGCGTG GCGGNGACGC ACGGCAAAAC 351 GACCACCGCG TCTATGCTCG CGTGGGTTTT GGAATATGCC GGACTCGCAC 401 CGGGCTTCNT TATCGGCGGC GTACCGGAAA ACTTCAGCGT TTCCGCCCGC 55 451 CTGCCGCAAA CGCCGCCCA AGACCCGAAC AGCCAATCGC CGTTTTTCGT 501 CATTGAAGCC GACGAATACG ACACCGCGTT TTTCGACAAA CGCTCCAAAT 551 TCGTGCATTA CCGTCCGCGT ACCGCCGTGT TGAACAATCT GGAATTCGAC 601 CACGCCGACA TCTTCGCCGA TTTGGGCGCG ATACAGACCC AGTTCCACCA 651 CCTCGTGCGT ACCGTGCCGT CTGAAGGCCT CATCGTCTGC AACGGCAGGCGC		101	AGATGTATCC	GCCGATGAGC	ACCCAGCTCG	AAGCCTTGGG	CATAGGCGTG
251 TGAACCGTGG GCTGCCTTAT ATTTCCGGCC CGCAATGGCT GGCTGAAAAC 301 NTGCTGCACC ATCATTGGNN ACTCGGCGTG GCGGNGACGC ACGGCAAAAC 351 GACCACCGCG TCTATGCTCG CGTGGGTTTT GGAATATGCC GGACTCGCAC 401 CGGGCTTCNT TATCGGCGGC GTACCGGAAA ACTTCAGCGT TTCCGCCCGC 55 451 CTGCCGCAAA CGCCGCGCCA AGACCCGAAC AGCCAATCGC CGTTTTTCGT 501 CATTGAAGCC GACGAATACG ACACCGCGTT TTTCGACAAA CGCTCCAAAT 551 TCGTGCATTA CCGTCCGCGT ACCGCCGTGT TGAACAATCT GGAATTCGAC 601 CACGCCGACA TCTTCGCCGA TTTGGGCGCG ATACAGACCC AGTTCCACCA 651 CCTCGTGCGT ACCGTGCCGT CTGAAGGCCT CATCGTCTGC AACGGACGGC		151	TATGAAGGCT	TCGACACCGC	GCAGTTGGAC	GAATTTAAAG	CCGACGTTTA
TGAACCGTGG GCTGCCTTAT ATTTCCGGCC CGCAATGGCT GGCTGAAAAC  301 NTGCTGCACC ATCATTGGNN ACTCGGCGTG GCGGNGACGC ACGGCAAAAC  351 GACCACCGCG TCTATGCTCG CGTGGGTTTT GGAATATGCC GGACTCGCAC  401 CGGGCTTCNT TATCGGCGGC GTACCGGAAA ACTTCAGCGT TTCCGCCCGC  451 CTGCCGCAAA CGCCGCGCCA AGACCCGAAC AGCCAATCGC CGTTTTTCGT  501 CATTGAAGCC GACGAATACG ACACCGCGTT TTTCGACAAA CGCTCCAAAT  551 TCGTGCATTA CCGTCCGCGT ACCGCCGTGT TGAACAATCT GGAATTCGAC  601 CACGCCGACA TCTTCGCCGA TTTGGGCGCG ATACAGACCC AGTTCCACCA  651 CCTCGTGCGT ACCGTGCCGT CTGAAGGCCT CATCGTCTGC	50	201	CGTTATCGGC	AATGTCGCCA	AGCGCGGGAT	GGATGTGGTT	GAAGCGATTT
55 GACCACCGCG TCTATGCTCG CGTGGGTTTT GGAATATGCC GGACTCGCAC 401 CGGGCTTCNT TATCGGCGGC GTACCGGAAA ACTTCAGCGT TTCCGCCCGC 55 451 CTGCCGCAAA CGCCGCGCCA AGACCCGAAC AGCCAATCGC CGTTTTTCGT 501 CATTGAAGCC GACGAATACG ACACCGCGTT TTTCGACAAA CGCTCCAAAT 551 TCGTGCATTA CCGTCCGCGT ACCGCCGTGT TGAACAATCT GGAATTCGAC 601 CACGCCGACA TCTTCGCCGA TTTGGGCGCG ATACAGACCC AGTTCCACCA 651 CCTCGTGCGT ACCGTGCCGT CTGAAGGCCT CATCGTCTGC AACGGACGGC		251	TGAACCGTGG	GCTGCCTTAT	ATTTCCGGCC	CGCAATGGCT	GGCTGAAAAC
55 451 CTGCCGCAAA CGCCGCCCA AGACCCGAAA ACTTCAGCGT TTCCGCCCGCCCAAC AGACCCGAAC AGCCAATCGC CGTTTTTCGT 501 CATTGAAGCC GACGAATACG ACACCGCGTT TTTCGACAAA CGCTCCAAAT 551 TCGTGCATTA CCGTCCGCGT ACCGCCGTGT TGAACAATCT GGAATTCGACCACAACCCGACA TCTTCGCCGA TTTTGGGCGCC ATACAGACCC AGTTCCACCAACCACCACCACCACCACCACCACCACCACCAC		301	NTGCTGCACC	ATCATTGGNN	ACTCGGCGTG	GCGGNGACGC	ACGGCAAAAC
55 451 CTGCCGCAAA CGCCGCGCCA AGACCCGAAC AGCCAATCGC CGTTTTTCGT 501 CATTGAAGCC GACGAATACG ACACCGCGTT TTTCGACAAA CGCTCCAAAT 551 TCGTGCATTA CCGTCCGCGT ACCGCCGTGT TGAACAATCT GGAATTCGAC 601 CACGCCGACA TCTTCGCCGA TTTGGGCGCG ATACAGACCC AGTTCCACCA 651 CCTCGTGCGT ACCGTGCCGT CTGAAGGCCT CATCGTCTGC AACGGACGGC		351	GACCACCGCG	TCTATGCTCG	CGTGGGTTTT	GGAATATGCC	GGACTCGCAC
501 CATTGAAGCC GACGAATACG ACACCGCGTT TTTCGACAAA CGCTCCAAAT 551 TCGTGCATTA CCGTCCGCGT ACCGCCGTGT TGAACAATCT GGAATTCGAC 601 CACGCCGACA TCTTCGCCGA TTTGGGCGCG ATACAGACCC AGTTCCACCA 651 CCTCGTGCGT ACCGTGCCGT CTGAAGGCCT CATCGTCTGC AACGGACGGC		401	CGGGCTTCNT	TATCGGCGGC	GTACCGGAAA		
501 CATTGAAGCC GACGAATACG ACACCGCGTT TTTCGACAAA CGCTCCAAAT 551 TCGTGCATTA CCGTCCGCGT ACCGCCGTGT TGAACAATCT GGAATTCGACCA 601 CACGCCGACA TCTTCGCCGA TTTGGGCGCG ATACAGACCC AGTTCCACCA 651 CCTCGTGCGT ACCGTGCCGT CTGAAGGCCT CATCGTCTGC AACGGACGGC	55	451	CTGCCGCAAA	CGCCGCGCCA	AGACCCGAAC		
601 CACGCCGACA TCTTCGCCGA TTTGGGCGCG ATACAGACCC AGTTCCACCA 651 CCTCGTGCGT ACCGTGCCGT CTGAAGGCCT CATCGTCTGC AACGGACGGC		501	CATTGAAGCC	GACGAATACG	ACACCGCGTT	TTTCGACAAA	CGCTCCAAAT
651 CCTCGTGCGT ACCGTGCCGT CTGAAGGCCT CATCGTCTGC AACGGACGGC		551	TCGTGCATTA	CCGTCCGCGT	ACCGCCGTGT	TGAACAATCT	GGAATTCGAC
		601	CACGCCGACA	TCTTCGCCGA	TTTGGGCGCG		
60 701 AGCAAAGCCT GCAAGACACT TTGGACAAAG GCTGCTGGAC GCCGGTGGAF		651	CCTCGTGCGT	ACCGTGCCGT	CTGAAGGCCT		
	60	701	AGCAAAGCCT	GCAAGACACT	TTGGACAAAG	GCTGCTGGAC	GCCGGTGGAA
751 AAATTCGGCA CGGAACACGG CTGGCAGGCC GGCGAAGCCA ATGCCGATGG	00		AAATTCGGCA	CGGAACACGG	CTGGCAGGCC	GGCGAAGCCA	ATGCCGATGG

		.,,
	801 CTCG	TTCGAC GTGTTGCTTG ACGGCAAAAA AGCCGGACAC GTCGCTTGGA
		GATGGG CGGACACAAC CGCATGAACG CGCTCGCNGT CATCGCCGCC
		GTCATG CCGGAGTNGA CATTCAGACG CCCTGCGAAG CCTTGAGCAC
	951 GTTT	AAAAAC GTCAAACGCC GCATGGAAAT CAAAGGCACG GCAAACGGTA
5		CGTTTA CGACGACTTC GCCCACCATC CGACCGCTAT CGAAACCACG
J		AAGGTT TGCGCCAGCG CGTCGGCGGC GCGCGCATCC TCGCCGTCCT
	· 1101 CGAA	CCGCGT TCCAATACGA TGAAGCTGGG TACGATGAAA GCCGCCCTGC
		AAGCCT CAAAGAAGCC GACCAAGTGT TCTGNTACGC CGGCGGCGCG
	1201 GACT	GGGACG TTGCCGAAGC CCTCGCGCCT TTGGGCGGCA GGCTGCACGT
10	1251 CGGC	AAAGAC TTCGATGCCT TCGTTGCCGA AATCGTGAAA AACGCCGAAG
10		CGACCA TATTTTGGTG ATGAGCAACG GCGGTTTCGG CGGAATACAC
		AACTGC TGGACGCTTT GAGATAG
	2002	
	This encodes a protein	n having amino acid sequence <seq 870="" id="">:</seq>
	1 mb chiedada a proton	
	1 MKHI	HIIGIG GTFMGGIAAI AKEAGFEXSG CDAKMYPPMS TQLEALGIGV
15		DTAQLD EFKADVYVIG NVAKRGMDVV EAILNRGLPY ISGPQWLAEN
15		HWXLGV AXTHGKTTTA SMLAWVLEYA GLAPGFXIGG VPENFSVSAR
	151 LPOT	PRODPN SOSPFFVIEA DEYDTAFFDK RSKFVHYRPR TAVLNNLEFD
	201 HADI	FADLGA IQTQFHHLVR TVPSEGLIVC NGRQQSLQDT LDKGCWTPVE
	251 KFGT	THOMAN TOTAL HAT DOWN OF HAM THOMAN TOWN TO THE TANK THE
20	301 ARHA	EHGWQA GEANADGSFD VLLDGKKAGH VAWSLMGGHN RMNALAVIAA
20	351 ARRA	GVDIQT ACEALSTFKN VKRRMEIKGT ANGITVYDDF AHHPTAIETT
	401 DWDV	RQRVGG ARILAVLEPR SNTMKLGTMK AALPASLKEA DQVFXYAGGA YAEALAP LGGRLHVGKD FDAFVAEIVK NAEAGDHILV MSNGGFGGIH
		ADALA* DALA*
	451 1855	DOUTE.
	ORF132a and ORF13	32-1 show 93.9% identity in 458 aa overlap:
25	orf132a.pep	MKHIHIIGIGGTFMGGIAAIAKEAGFEXSGCDAKMYPPMSTQLEALGIGVYEGFDTAQLD
23	OIII32a.pep	
	orf132-1	MKHIHIIGIGGTFMGGLAAIAKEAGFEVSGCDAKMYPPMSTQLEALGIDVYEGFDAAQLD
	022232	THE THE PARTY OF T
	orf132a.pep	EFKADVYVIGNVAKRGMDVVEAILNRGLPYISGPQWLAENXLHHHWXLGVAXTHGKTTTA
30	OZZZZOBUJĘCĘ	
•	orf132-1	EFKADVYVIGNVAKRGMDVVEAILNLGLPYISGPQWLSENVLHHHWVLGVAGTHGKTTTA
	<b>V</b>	The state of the s
	orf132a.pep	SMLAWVLEYAGLAPGFXIGGVPENFSVSARLPQTPRQDPNSQSPFFVIEADEYDTAFFDK
	• ·•	
35	orf132-1	SMLAWVLEYAGLAPGFLIGGVPENFGVSARLPQTPRQDPNSQSPFFVIEADEYDTAFFDK
		The state of the s
	orf132a.pep	RSKFVHYRPRTAVLNNLEFDHADIFADLGAIQTQFHHLVRTVPSEGLIVCNGRQQSLQDT
	orf132-1	RSKFVHYRPRTAVLNNLEFDHADIFADLGAIQTQFHYLVRTVPSEGLIVCNGRQQSLQDT
40		
	orf132a.pep	LDKGCWTPVEKFGTEHGWQAGEANADGSFDVLLDGKKAGHVAWSLMGGHNRMNALAVIAA
	orf132-1	LDKGCWTPVEKFGTEHGWQAGEANADGSFDVLLDGKTAGRVKWDLMGRHNRMNALAVIAA
		· · · · · · · · · · · · · · · · · · ·
45	orf132a.pep	ARHAGVDIQTACEALSTFKNVKRRMEIKGTANGITVYDDFAHHPTAIETTIQGLRQRVGG
		- 111:111111111111::1111111111111111111
	orf132-1	ARHVGVDIQTACEALGAFKNVKRRMEIKGTANGITVYDDFAHHPTAIETTIQGLRQRVGG
	orf132a.pep	ARILAVLEPRSNTMKLGTMKAALPASLKEADQVFXYAGGADWDVAEALAPLGGRLHVGKD
50		
	orf132-1	ARILAVLEPRSNTMKLGTMKSALPVSLKEADQVFCYAGGVDWDVAEALAPLGGRLNVGKD

# Homology with a predicted ORF from N.gonorrhoeae

orf132a.pep

orf132-1

55

ORF132 shows 89.6% identity over 259 aa overlap with a predicted ORF (ORF132ng) from N. gonorrhoeae:

60	orf132.pep	MKHIHIIGIGGTFMGGLAAIAKEAGFEVSGCDAKMYPPMSTQLEALGIDVYEGFDAAQLD	60
	orfl32ng	MKHIHIIGIGGTFMGGIAAIAKEAGFKVSGCDAKMYPPMSTQLEALGIGVHEGFDAAOLE	60

PCT/IB98/01665

-471-

		.: 1 CDO TD 0	715 horring	- amir
15	orf132ng		261	
	orf132.pep	FGQRLLDAGGKIRHGTRLA	259	
10	orf132ng	TLQIRALSPAYRRVEQSGIRPRR	HLRRLGRDTDPVPPPRAHRTIRRPHRLQRTAAKPARY	240
10	orf132.pep	TFXIRALPSAYRRVEQSGIRPRR	HLCRLGRDTDPVPLPRAYRAVXRLNRLQRTAAKPARY	240
	orf132ng	SMLAWVLEYAGLAPGFLIGGVPG	KFRRFRPPTANAASRPEQQIAVFRHRSRRIRHRLFRQ	180
5	orf132.pep	SMLAWVLEYAGLAPGFLIGGVXG	KFRRFRPPAANAAPRPEQPIAVFRHRSRRIRHRLFRQ	180
	orf132ng	EFQADIYVIGNVARRGMDVVEAI	LNRGLPYISGPQWLAENVLHHHWVLGVAGTHGKTTTA	120
	orf132.pep	EFKADVYVIGNVAKRGMDVVEAI	LNLGLPYISGPQWLSENVLHHHWVLGVAGTHGKTTTA	120

An ORF132ng nucleotide sequence <SEQ ID 871> was predicted to encode a protein having amino acid sequence <SEQ ID 872>:

```
MKHIHIIGIG GTFMGGIAAI AKEAGFKVSG CDAKMYPPMS TQLEALGIGV
                     HEGFDAAQLE EFQADIYVIG NVARRGMDVV EAILNRGLPY ISGPQWLAEN
                     VLHHHWVLGV AGTHGKTTTA SMLAWVLEYA GLAPGFLIGG VPGKFRRFRP
20
                101
                     PTANAASRPE QQIAVFRHRS RRIRHRLFRQ TLQIRALSPA YRRVEQSGIR
                151
                     PRRHLRRLGR DTDPVPPPRA HRTIRRPHRL QRTAAKPARY FGQRLLDAGG
                201
                251 KIRHRTRLAD W*
```

Further work revealed the following gonococcal DNA sequence <SEQ ID 873>:

```
ATGAAACACA TCCACATTAT CGGTATCGGC GGCACGTTTA TGGGCGGGAT
25
                    TGCCGCCATT GCCAAAGAAG CCGGGTTCAA AGTCAGCGGT TGCGACGCGA
                 51
                     AGATGTATCC GCCGATGAGC ACCCAGCTCG AAGCCTTGGG CATAGGCGTA
                101
                     CACGAAGGCT TCGATGCCGC GCAGTTGGAA GAATTTCAAG CCGATATTTA
                151
                    CGTCATCGGC AATGTCGCCA GGCGCGGGAT GGATGTGGTC GAGGCGATTT
                201
                     TGAACCGTGG GCTGCCTTAT ATTTCCGGCC CGCAATGGCT GGCTGAAAac
30
                251
                     GTGCtgcacc atcaTTGGgt ACTCGGCGTG GcagggaCGC ACGGcaaAac
                301
                     gaccaCcGcg tCCATGCTCG CCTGGGTCTT GGAATATGCC GGACTCGCGC
                351
                     CGGGCTTCCT CATCGGCGGt gtaccggaAA ATTTCGGCGT TTCCGCCCGC
                401
                     CTACCGCAAA CGCCGCGTCA AGACCCGAAC AGCAAATCGC CGTTTTTCGT
                451
                     CATCGAAGCC GACGAATACG ACACCGCCTT TTTCGACAAA CGCTCCAAAT
35
                501
                     TCGTGCATTA TCGCCCGCGT ACCGCCGTGT TGAACAATCT GGAATTCGAC
                551
                     CACGCCGACA TCTTCGCCGA CTTGGGCGCG ATACAGACCC AGTTCCACCA
                601
                     CCTCGTGCGC ACCGTACCAT CCGAAGGCCT CATCGTCTGC AACGGACAGC
                651
                     AGCAAAGCCT GCAAGATACT TTGGACAAAG GCTGCTGGAC GCCGGTGGAA
                701
                    AAATTCGGCA CCGGACACGG CTGGCAGATT GGTGAAGTCA ATGCCGACGG
40
                751
                     CTCGTTCGAC GTATTGCTTG ACGGCAAAAA AGCCGGACAC GTCGCATGGG
                801
                    ATTTGATGGG CGGACACAAC CGCATGAACG CGCTCGCCGT CATCGCTGCC
                851
                901 GCACGCCATG CCGGAGTCGA TGTTCAGACG GCCTGCGAAG CCTTGGGTGC
                     GTTTAAAAAC GTCAAACGCC GCATGGAAAT CAAAGGCACG GCAAACGGCA
                951
                     TCACCGTTTA CGACGATTTC GCCCACCACC CGACCGCCAT CGAAACCACG
45
               1001
               1051 ATTCAAGGTT TGCGCCAACG TGTCGGCGGC GCGCGCATCC TCGCCGTCCT
                     CGAGCCGCGT TCCAACACCA TGAAACTCGG CACGATGAAG TCCGCCCTGC
               1101
                     CCGCAAGCCT CAAAGAAGCC GACCAAGTGT TCTGCTACGC CGGCGGCGCG
               1151
                     GACTGGGACG TTGCCGAAGC CCTCGCGCCT TTGGGCTGCA GGCTGCGCGT
               1201
                     CGGTAAAGAT TTCGATACCT TCGTTGCCGA AATTGTGAAA AACGCCCGAA
 50
               1251
                     CCGGCGACCA TATTTTGGTG ATGAGCAACG GCGGTTTCGG CGGAATACAC
               1301
               1351 ACCAAACTGC TGGACGCTTT GAGATAG
```

# This corresponds to the amino acid sequence <SEQ ID 874; ORF132ng-1>:

	1	MKHIHIIGIG	GTFMGGIAAI	AKEAGFKVSG	CDAKMYPPMS	TQLEALGIGV
55	51	HEGFDAAQLE	EFQADIYVIG	NVARRGMDVV	EAILNRGLPY	ISGPQWLAEN
	101	VLHHHWVLGV	AGTHGKTTTA	SMLAWVLEYA	GLAPGFLIGG	VPENFGVSAR
	151				RSKFVHYRPR	
	201	HADIFADLGA	IQTQFHHLVR	TVPSEGLIVC	NGQQQSLQDT	LDKGCWTPVE
	251	KFGTGHGWQI	GEVNADGSFD	VLLDGKKAGH	VAWDLMGGHN	RMNALAVIAA
60	301	ARHAGVDVQT	ACEALGAFKN	VKRRMEIKGT	ANGITVYDDF	AHHPTAIETT
	351	IQGLRQRVGG	ARILAVLEPR	SNTMKLGTMK	SALPASLKEA	DOVECYAGGA
	401	DWDVAEALAP	LGCRLRVGKD	FDTFVAEIVK	NARTGDHILV	MSNGGFGGIH
	451	TKLLDALR*				

# ORF132ng-1 and ORF132-1 show 93.2% identity in 458 aa overlap:

	orf132ng-1.pep orf132-1	MKHIHIIGIGGTFMGGIAAIAKEAGFKVSGCDAKMYPPMSTQLEALGIGVHEGFDAAQ 	1:
5	orf132ng-1.per	> EFQADIYVIGNVARRGMDVVEAILNRGLPYISGPQWLAENVLHHHWVLGVAGTHGKTI	TA
	orf132-1	:  :       :	III TTA
10	orf132ng-1.pep	SMLAWVLEYAGLAPGFLIGGVPENFGVSARLPQTPRQDPNSKSPFFVIEADEYDTAFE	DK
	orf132-1		PDK
15	orf132ng-1.pep	RSKFVHYRPRTAVLNNLEFDHADIFADLGAIQTQFHHLVRTVPSEGLIVCNGQQQSLQ	QDT
1.5	orf132-1	RSKFVHYRPRTAVLNNLEFDHADIFADLGAIQTQFHYLVRTVPSEGLIVCNGRQQSLQ	
	orf132ng-1.pep	LDKGCWTPVEKFGTGHGWQIGEVNADGSFDVLLDGKKAGHVAWDLMGGHNRMNALAVI	EAA .
20	orf132-1	LDKGCWTPVEKFGTEHGWQAGEANADGSFDVLLDGKTAGRVKWDLMGRHNRMNALAV	IAA.
	orf132ng-1.per	ARHAGVDVQTACEALGAFKNVKRRMEIKGTANGITVYDDFAHHPTAIETTIQGLRQRV	/GG i I I
25	orf132-1	ARHVGVDIQTACEALGAFKNVKRRMEIKGTANGITVYDDFAHHPTAIETTIQGLRQR\	<b>V</b> GG
		ARILAVLEPRSNTMKLGTMKSALPASLKEADQVFCYAGGADWDVAEALAPLGCRLRVC	111
20	orf132-1	ARILAVLEPRSNTMKLGTMKSALPVSLKEADQVFCYAGGVDWDVAEALAPLGGRLNV	GKD
30		FDTFVAEIVKNARTGDHILVMSNGGFGGIHTKLLDALRX	
	orf132-1 In addition ORF132	rdafvaeivknaevgdhilvmsnggfggihgkllealrx  ng-1 is homologous to a hypothetical E.coli protein:	٠
		•	
35	ORF_0457 (Eschin fbp-pmba in Score = 474	<pre>ypothetical protein o457 - Escherichia coli &gt;gi 537075 (U140) herichia coli] &gt;gi 1790680 (AE000494) hypothetical 48.5 kD period region [Escherichia coli] Length = 457 bits (1207), Expect = e-133 249/439 (56%), Positives = 294/439 (66%), Gaps = 13/439 (25%)</pre>	protein
40		AGFKVSGCDAKMYPPMSTQLEALGIGVHEGFDAAQLEEFQADIYVIGNVARRGMDVVE	
		G +V+G DA +YPPMST LE GI + +G+DA+QLE Q D+ +IGN RG VE LGHEVTGSDANVYPPMSTLLEKQGIELIQGYDASQLEP-QPDLVIIGNAMTRGNPCVE	79
4.5	Query: 82 AI	LNRGLPYISGPQWLAENVLHHHWVLGVAGTHGKTTTASMLAWVLEYAGLAPGFLIGGV	141
45		L + +PY+SGPQWL + VL WVL VAGTHGKTTTA M W+LE G PGF+IGGV LEKNIPYMSGPQWLHDFVLRDRWVLAVAGTHGKTTTAGMATWILEQCGYKPGFVIGGV	139
		NFGVSARLPQTPRQDPNSKSPFFVIEADEYDTAFFDKRSKFVHYRPRTAVLNNLEFDH	201
50		NF VSA L +S FFVIEADEYD AFFDKRSKFVHY PRT +LNNLEFDH NFEVSAHLGESDFFVIEADEYDCAFFDKRSKFVHYCPRTLILNNLEFDH	190
	Query: 202 AD	IFADLGAIQTQFHHLVRTVPSEGLIVCNGQQQSLQDTLDKGCWTPVEKFGTGHGWQIG : IF DL AIQ QFHHLVR VP +G I+ +L+ T+ GCW+ E G WQ	261
55	Sbjct: 191 AD	IF DL AIQ QFHHLVR VP +G I+ +L+ T+ GCW+ E G WQ IFDDLKAIQKQFHHLVRIVPGQGRIIWPENDINLKQTMAMGCWSEQELVGEQGHWQAK	250
		NADGS-FDVLLDGKKAGHVAWDLMGGHNRMNALAVIAAARHAGVDVQTACEALGAFKN DS++VLLDG+KGVWL+GHNNLIAAARHGVAALG+FN	320
		TTDASEWEVLLDGEKVGEVKWSLVGEHNMHNGLMAIAAARHVGVAPADAANALGSFIN	310
60	Query: 321 VK	RRMEIKGTANGITVYDDFAHHPTAIETTIQGLRQRVGG-ARILAVLEPRSNTMKLGTM RR+E++G ANG+TVYDDFAHHPTAI T+ LR +VGG ARI+AVLEPRSNTMK+G	379
	Sbjct: 311 AR	RRLELRGEANGVTVYDDFAHHPTAILATLAALRGKVGGTARIIAVLEPRSNTMKMGIC	
65	K	ALPASLKEADQVF-CYAGGADWDVAEALAPLGCRLRVGKDFDTFVAEIVKNARTGDHI L SL AD+VF W VAE D DT +VK A+ GDHI	
	•	DLAPSLGRADEVFLLQPAHIPWQVAEVAEACVQPAHWSGDVDTLADMVVKTAQPGDHI	430
	Query: 439 LV	MSNGGFGGIHTKLLDAL 457	

LVMSNGGFGGIH KLLD L Sbjct: 431 LVMSNGGFGGIHQKLLDGL 449

Based on this analysis, it was predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF132-1 (26.4kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 20A shows the results of affinity purification of the His-fusion protein, and Figure 20B shows the results of expression of the GST-fusion in *E.coli*. Purified His-fusion protein was used to immunise mice, whose sera were used for FACS analysis (Figure 20C) and ELISA (positive result). These experiments confirm that ORF132 is a surface-exposed protein, and that it is a useful immunogen.

Example 103

5

10

The following partial DNA sequence was identified in N.meningitidis <SEQ ID 875>

```
...CCGGGCTATT ACGGCTCGGA TGACGAATTT AAGCGGGCAT TCGGAGAAAA
                       CTCGCCGACA TmCAAGAAAC ATTGCAACCG GAGCTGCGGG ATTTATGAAC
                51
15
                       CCGTATTGAA AAAATACGGC AAAAAGCGCG CCAACAACCA TTCGGTCAGC
                101
                       ATTAGTGCGG ACTTCGGCGA TTATTTCATG CCGTTCGCCA GCTATTCGCG
                151
                       CACACACCGT ATGCCCAACA TCCAAGAAAT GTATTTTTCC CAAATCGGCG
                201
                       ACTCCGGCGT TCACACCGCC TTAAAACCAG AGCGCGCAAA CACTTGGCAA
                251
                       TTTGGCTTCr ATACCTATAA AAAAGGATTG TTAAAACAAG ATGATACATT
                301
                       AGGATTAAAA CTGGTCGGCT ACCGCAGCCG CATCGACAAC TACATCCACA
20
                351
                       ACGTTTACGG GAAATGGTGG GATTTGAACG GGGATATTCC GAGCTGGGTC
                401
                       AGCAGCACCG GGCTTGCCTA CACCATCCAA CATCGCTATT TCAWAGACAA
                451
                       AGTGCATCAA nnnnnnnnn nnnnnnnnn nnnntacgat tatgggcgtt
                501
                       TTTTCACCAA CCTTTCTTAC GCCTATCAAA AAAGCACGCA ACCGACCAAC
                551
25
                601
                       TTCAGCGATG CGAGCGAATC GCCCAACAAT GCGTCCAAAG AAGACCAACT
                       CAAACAAGGT TATGGGTTGA GCAGGGTTTC CGCCCTGCCG CGAGATTACG
                651
                       GACGTTTGGA AGTCGGTACG CGCTGGTTGG GCAACAAACT GACTTTGGGC
                701
                       GGCGCGATGC GCTATTTCGG CAAGAGCATC CGCGCGACGG CTGAAGAACG
                751
                       CTATATCGAC GGCACCAACG GGGGAAATAC CAGCAATTTC CGGCAACTGG
                801
                       GCAAGCGTTC CATCAAACAA ACCGAAACTC TTGCCCGCCA GCCTTTGATT
30
                851
                       TTWGATTTTA ACGCCGCTTA CGAGCCGAAG AAAAACCTTA TTTTCCGCGC
                901
                       CGAAGTCAAA AATCTGTTCG ACAGGCGTTA TATCGATCCG CTCGATGCGG
                951
                       GCAATGATGC GGCAAC.GAG CGTTATTACA GCTCGTTCGA CCCGAAAGAC
               1001
                       AAGGACTTAG ACGTAACGTG TAATGCTGAT AAAACGTTGT GCAACGGCAA
               1051
                       ATACGGCGGC ACAAGCAAAA GCGTATTGAC CAATTTTGCA CGCGGACGCA
35
               1101
                       CCTTTTTgAT GACGATGAGC TACAAGTTTT AA
               1151
```

This corresponds to the amino acid sequence <SEQ ID 876; ORF133>:

```
.. PGYYGSDDEF KRAFGENSPT XKKHCNRSCG IYEPVLKKYG KKRANNHSVS
                       ISADFGDYFM PFASYSRTHR MPNIQEMYFS QIGDSGVHTA LKPERANTWQ
                51
                       FGFXTYKKGL LKQDDTLGLK LVGYRSRIDN YIHNVYGKWW DLNGDIPSWV
40
                101
                       SSTGLAYTIQ HRXFXDKVHQ XXXXXXXXYD YGRFFTNLSY AYQKSTQPTN
                151
                       FSDASESPNN ASKEDQLKQG YGLSRVSALP RDYGRLEVGT RWLGNKLTLG
                201
                       GAMRYFGKSI RATAEERYID GTNGGNTSNF RQLGKRSIKQ TETLARQPLI
                251
                       XDFNAAYEPK KNLIFRAEVK NLFDRRYIDP LDAGNDAAXE RYYSSFDPKD
                301
                       KDXDVTCNAD KTLCNGKYGG TSKSVLTNFA RGRTFLMTMS YKF*
45
```

Further work revealed the further partial DNA sequence <SEQ ID 877>:

50	51 101 151 201	GAGGCGCAGA ACCGAAAGAC AGGATATATT CCCGGTGCGT TATTCGCGGC	AAAAAGTGT CAAATCCAGC TTACACAGCA GACAGCGGGT	TTACCGATGC GAAAACCTCG AGATAAAAGC TCGGGCGGGT	GCGTGCCGTA ACAACATCGT TCGGGCATTG CAATACGATG	TCGACCCGTC ACGCAGCATC TGTCTTTGAA GTGGACGGCA
	251	TCACGCAGAC	CTTTTATTCG	ACTTCTACCG	ATGCGGGCAG	GGCAGGCGGT

	301	TCATCTCAAT	TCGGTGCATC	TGTCGACAGC	AATTTTATTG	CCGGACTGGA
	351	TGTCGTCAAA	GGCAGCTTCA	GCGGCTCGGC	AGGCATCAAC	AGCCTTGCCG
	401	GTTCGGCGAA	TCTGCGGACT	TTAGGCGTGG	ATGACGTCGT	TCAGGGCAAT
	451	AATACCTACG	GCCTGCTGCT	AAAAGGTCTG	ACCGGCACCA	ATTCAACCAA
5				TAGGTGCGCG		
,	551			GGGCACAGCA		
	· 601	TACCGCGTGG	GCGGCGGCGG	GCAGCACATC	CCAAATTTTC	CCCCCCAATA
	651	THUCCOCOTOG	CCCAACCACC	GATATTTTGT	ACAACACCCT	CCGCGGGAAIA
10	701			TGGGAGCGGG		
10	751			TTACAACAAC		
	801	CGAAGAGCAT	GACAAAAGCT	GGCGGGAAAA	CCTg.CaCCG	CAATACGACA
	851			AGCCTGAAGC		
	901			CGTATTCAAT		
	951			GCAGCCGCAA		
15	1001	AGTTCAATTA	CGGTTTGTCT	TTGAACCCGT	ATACCAACCT	CAATCTGACC
	1051			GCAGAAATAT		
	1101	AGGCTGGGGG	CTTTTAAAGG	ATTTTGAAAC	CTACAACAAC	GCGAAAATCC
	1151	TCGACCTCAA	CAACACCGCC	ACCTTCCGGC	TGCCCCGCGA	AACCGAGTTG
	1201	CAAACCACTT	TGGGCTTCAA	TTATTTCCAC	DACCADTACG	CODDADACCC
20	1251			TGTTTTTCGA		
20	1301			CGGTTTAAGG		
	1351					
	1351			ACCGGCCGGC		
	1401			AAAAAGACAT		
0.5	1451			TTCGGCGGCG		
25	1501			GGCATTCGGA		
	1551			GCGGGATTTA		
	1601			AACCATTCGG		
	1651			CGCCAGCTAT		
	1701 1751 1801	CAACATCCAA	GAAATGTATT	TTTCCCAAAT	CGGCGACTCC	GGCGTTCACA
30	1751			GCAAACACTT		
	1801	TATAAAAAAG	GATTGTTAAA	ACAAGATGAT	ACATTAGGAT	TAAAACTGGT
	1851			ACAACTACAT		
	1901			ATTCCGAGCT		
	1951			CAATTTCAAA		
35	2001			ACGATTATGG		
55	2051			ACGCAACCGA		
	2101			CAAAGAAGAC		
	2151			TGCCGCGAGA		
40	2201			AAACTGACTT		
40	2251			GACGGCTGAA		
	2301			ATTTCCGGCA		
	2351			CGCCAGCCTT		
	2401	GCTTACGAGC	CGAAGAAAAA	CCTTATTTTC	CGCGCCGAAG	TCAAAAATCT
	2451			ATCCGCTCGA		
45	2501	CGCAGCGTTA	TTACAGCTCG	TTCGACCCGA	AAGACAAGGA	CGAAGACGTA
	2551	ACGTGTAATG	CTGATAAAAC	GTTGTGCAAC	GGCAAATACG	GCGGCACAAG
	2601	CAAAAGCGTA	TTGACCAATT	TTGCACGCGG	ACGCACCTTT	TTGATGACGA
	2651	TGAGCTACAA				
	This correspond	s to the amin	o acid seque	ence <seq i<="" th=""><th>D 878; ORF</th><th>133-1&gt;:</th></seq>	D 878; ORF	133-1>:
50	. 1	EACTOVIEDY	HAKVKBADKU	KKVFTDARAV	STRONTERCO	ENTINATABET
30	51			DSGFGRVNTM		
	101			GSFSGSAGIN		
	151			MAAIGARKWL		
E E	201			RKORYFVOEG		
55	251			DKSWRENLXP		
	301			TKIGSRKIIN		
	351	AAYNSGRQKY	PKGSKFTGWG	LLKDFETYNN	AKILDLNNTA	TFRLPRETEI
	401	QTTLGFNYFH	NEYGKNRFPE	ELGLFFDGPD	QDNGLYSYLG	RFKGDKGLLE
	451	QKSTIVQPAG	SQYFNTFYFD	AALKKDIYRL	NYSTNTVGYR	FGGEYTGYY
60	501	SDDEFKRAFG	ENSPTYKKHO	NRSCGIYEPV	LKKYGKKRAN	NHSVSISADE
	551			EMYFSQIGDS		
	601			SRIDNYIHNV		
	651			ELNYDYGRFF		
	701			VSALPRDYGR		
65	701 751					
U)				NTSNFRQLGK		
	801			RYIDPLDAGN		t DEVNVOED/
	851	TUNADKTLON	GKIGGTSKSV	LINFARGRIE	PWIW21KF.	•

Computer analysis of this amino acid sequence gave the following results:

# Homology with with the probable TonB-dependent receptor HI121 of *H.influenzae* (accession number U32801) ORF133 and HI121 show 57% aa identity in 363aa overlap:

	Orf133:	31	IYEPVLKKYGKKRANNHSVSISADFGDYFMPFASYSRTHRMPNIQEMYFSQIGDSGVHTA I EP+L K G K+A NHS ++SA+ DYFMPF +YSRTHRMPNIQEM+FSQ+ ++GV+TA	90
5	HI121:	563	INEPILHKSGHKKAFNHSATLSAELSDYFMPFFTYSRTHRMPNIQEMFFSQVSNAGVNTA	622
	Orf133:	91	LKPERANTWQFGFXTYKKGLLKQDDTLGLKLVGYRSRIDNYIHNVYGKWWDLNGDIPSWV LKPE+++T+O GF TYKKGL ODD LG+KLVGYRS I NYIHNVYG WW +P+W	150
10	HI121:	623	LKPEQSDTYQLGFNTYKKGLFTQDDVLGVKLVGYRSFIKNYIHNVYGVWWRDGMPTWA	680
10	Orf133:	151	SSTGLAYTIQHRXFXDKVHXXXXXXXXYDYGRFFTNLSYAYQKSTQPTNFSDASESPNN SG YTI H+ + V YD GRFF N+SYAYQ++ QPTN++DAS PNN	210
	HI121:	681	ESNGFKYTIAHQNYKPIVKKSGVELEINYDMGRFFANVSYAYQRTNQPTNYADASPRPNN	740
15	Orf133:	211	ASKEDQLKQGYGLSRVSALPRDYGRLEVGTRWLGNKLTLGGAMRYFGKSIRATAEERYID AS+ED LKQGYGLSRVS LP+DYGRLE+GTRW KLTLG A RY+GKS RAT EE YI+	270
	HI121:	741	ASQEDILKQGYGLSRVSMLPKDYGRLELGTRWFDQKLTLGLAARYYGKSKRATIEEEYIN	800
20	Orf133:	271	GTNGGNTSNFRQLGKRSIKQTETLARQPLIXDFNAAYEPKKNLIFRAEVKNLFDRRYIDP G+ + R+ ++K+TE + +OP+I D + +YEP K+LI +AEV+NL D+RY+DP	330
20	HI121:	801	GSR-FKKNTLRRENYYAVKKTEDIKKQPIILDLHVSYEPIKDLIIKAEVQNLLDKRYVDP	859
	Orf133:	331	LDAGNDAAXERYYSSFDPKDKDXDVTCNADKTLCNGKYGGTSKSVLTNFARGRTFLMTMS LDAGNDAA +RYYSS + + C D + C GG+ K+VL NFARGRT++++++	390
25	HI121:	860	LDAGNDAASQRYYSSLNNSIECAQDSSACGGSDKTVLYNFARGRTYILSLN	910
	Orf133:	391	YKF 393 YKF	
30	HI121:	911	YKF 913	
.30				

## Homology with a predicted ORF from N. meningitidis (strain A)

ORF133 shows 90.8% identity over a 392aa overlap with an ORF (ORF133a) from strain A of N. meningitidis:

						10	20	30
35	orf133.pep				PGYYGS	DDEFKRAFGE	NSPTXKKHCN:	RSCGI
					111 1	1111111111	1111 1111	:1111
	orf133a	FYFDAAI	LKKDIYRLN'	YSTNTVGYRF	GGXYTGYYXS	DDEFKRAFGEI	SPTYXKHCN	QSCGI
		450	460	470	480	490	500	
40			40	50	60	70	80	90
	orf133.pep	YEPVLKI	KYGKKRANN	HSVSISADFG	DYFMPFASYS	RTHRMPNIQE	<b>MYFSQIGDSG</b>	VHTAL
		111111	11111111	11111111111	1111111111		111111111	1111
	orf133a	YEPVLKI	KYGKKRANN.	HSVSISADFG	DYFMPFASYS	RTHRMPNIQE	MYFSQIGDSG	VHTAL
		510	520	530	540	550	560	
45								
			100	110	120	130	140	150
	orf133.pep	KPERAN'	IWQFGFXTY	KKGLLKQDDT	LGLKLVGYRS	RIDNYIHNVY	GKWWDLNGDI	PSWVS
	-	111111	111111 11	111111111	1111111111	111 111111	11111111:1	11111
	orf133a	KPERAN'	TWQFGFNTY.	KKGLLKQDDI	LGLKLVGYRS	RIDXYIHNVY	GKWWDLNGNI	PSWVS
50		570	580	590	600	610	620	
		;	160	170	180	190	200	210
	orf133.pep	STGLAY	TIQHRXFXD	KVHQXXXXXX	XXYDYGRFFT	NLSYAYQKST	OPTNESDASE	SPNNA
		111111	11111 1 1	111:	111 1111	111111111111		11111
55	orf133a	STGLAY	TIQHRNFKD	KVHKHGFELE	LNYDYXRFFI	NLSYAYOKST	OPTNFSDASE	SPNNA
		630	640	650	660	670	680	
		:	220	230	240	250	260	270
	orf133.pep	SKEDOL	KOGYGLSRV	SALPRDYGRL	EVGTRWLGNK	LTLGGAMRYF	GKSIRATAEE	RYIDG
60		111111	пини	11111111111	11111111111	11111111111	111111111111	HHI
* *	orf133a	SKEDOL	KOGYGLSRV	SALPRDYGRI	EVGTRWLGNE	LTLGGAMRYF	GKSIRATAEE	RYIDX
		690	700	710	720	730	740	
			280	290	300	310	320	330
65	orf133.pep	TNGGNT	SNFRQLGKR	SIKOTETLAR	ROPLIXDENA	YEPKKNLIFR	AEVKNLFDRR	YIDPL
			- · · <del>-</del>					

```
TNGXXTSNFRQLGKRSIXQTETLARQPLIFDXYAAYEPKKXLIFRAEVKNLFDRRYIDPL
          orf133a
                     750
                               760
                                         770
                                                   780
                                                              790
                                                                        800
 5
                             340
                                        350
                                                  360
                                                            370
                                                                      380
                       DAGNDAAXERYYSSFDPKDKDXDVTCNADKTLCNGKYGGTSKSVLTNFARGRTFLMTMSY
          orf133.pep
                        DAGNDAATQRYYSSFDPKDKDEEVTCNDDNTLCNGKYGGTSKSVLTNFARGXTFLITMSY
          orf133a
                     810
                               820
                                          830
                                                    840
                                                              850
10
                        KFX
          orf133.pep
                        111
          orf133a
                        KFX
15
     A partial ORF133a nucleotide sequence <SEQ ID 879> is:
                    AAAGACAAAA AAGTGTTTAC CGATGCGCGT GCCGTATCGA CCCGTCAGGA
                     TATATTCAAA TCCANCGAAA ACCTCGACAA CATCGTACGC ANCATCCCCG
                 51
                     GTGCGTTTAC ACANCAANAT AAAAGCTCGG GCNTTGTGTC TTTGAATATT
               101
20
                     CGCNGCGACA GCGGGTTCGG GCGGGTCAAT ACNATGGTNG ACGGCATCAC
               151
                     NCANACCTTT TATTCGACTT CTACCGATGC GGGCAGGGCA GGCGGTTCAT
                201
                     CTCAATTCGG TGCATCTGTC GACAGCAATT TTATNGCCGG ACTGGATGTC
               251
                     GTCAAAGGCA GCTTCAGCGG CTCGGCAGGC ATCAACAGCC TTGCCGGTTC
                301
                     GGCGAATCTG CGGACTTTAN GCGTGGATGA TGTCGTTCAG GGCAATANTA
                351
25
                401
                     CNTACGGCCT GCTGCTAAAA GGTCTGACCG GCACCAATTC AACCAAAGGT
                     AATGCGATGG CGGCGATAGG TGCGCGCAAA TGGCTGGAAA GCGGAGCATC
                451
                     TGTCGGTGTG CTTTACGGGC ACAGCAGGCG CAGCGTGGCG CAAAATTACC
                501
                     GCGTGGCGG CGCCGGCAG CACATCGGAA ATTTTGGCGC GGAATATCTG
                551
                     GAACGACGCA AGCAACGATA TTTTGAGCAA GAAGGCGGGT TGAAATTCAA
                601
30
                651
                     TTCCAACAGC GGAAAATGGG AGCGGGATTT CCAAAAGTCG TACTGGAAAA
                     CCAAGTGGTA TCAAAAATAC GATGCCCCCC AAGAACTGCA AAAATACATC
                701
                751
                     GAAGGTCATG ATAAAAGCTG GCGGGAAAAC CTGGCGCCGC AATACGACAT
                     CACCCCATC GATCCGTCCA GCCTGAAGCN GCAGTCGGCA GGCAACCTGT
                801
                851
                     TTAAATTGGA ATACGACGGC GTATTCAATA AATACACGGC GCAATTTCGC
35
                     GATTTAAACA CCAAAATCGG CAGCCGCAAA ATCATCAACC GCAATTATCA
                901
                     ATTCAATTAC GGTTTGTCTT TGAACCCGTA TACCAACCTC AATCTGACCG
                951
                     CAGCCTACAA TTCGGGCAGG CAGAAATATC CGAAAGGGTC GAAGTTTACA
               1001
                     GGCTGGGGGC TTTTNAAAGA TTTTGAAACC TACAACAACG CAAAAATCCT
               1051
                     CGACCTCANC AACACCTCCA CCTTCCGGCT GCCCCGTGAA ACCGAGTTGC
               1101
                     AAACCACTTT GGGCTTCAAT TATTTCCACA ACGAATACGG CAAAAACCGC
TTTCCTGAAG AATTGGGGCT GTTTTTCGAC GGTCCGGATC ANGACAACGG
40
               1151
               1201
               1251
                     GCTTTATTCC TATTTGGGGC GGTTTAAGGG CGATAAAGGG CTGCTGCCCC
                     AAAAATCAAC CATTGTCCAA CCGGCCGGCA GCCAATATTT CAACACGTTC
               1301
                     TACTTCGATG CCGCGCTCAA AAAAGACATT TACCGCTTAA ACTACAGCAC
               1351
                     CAATACCGTC GGCTACCGTT TCGGCGGCNA ATATACGGGC TATTACNGCT
 45
               1401
                     CGGATGACGA ATTTAAGCGG GCATTCGGAG AAAACTCGCC GACATACANG
               1451
                     AAACATTGCA ACCAGAGCTG CGGAATTTAT GAACCCGTAT TGAAAAAATA
               1501
                     CGGCAAAAAG CGCGCCAACA ACCATTCGGT CAGCATTAGT GCGGACTTCG
               1551
                     GCGATTATTT CATGCCGTTC GCCAGCTATT CGCGCACACA CCGTATGCCC
               1601
 50
                     AACATCCAAG AAATGTATTT TTCCCAAATC GGCGACTCCG GCGTTCACAC
               1651
                     CGCCTTAAAA CCAGAGCGCG CAAACACTTG GCAATTTGGC TTCAATACCT
               1701
               1751
                     ATAAAAAGG ATTGTTAAAA CAAGATGATA TATTAGGATT AAAACTGGTC
                     GGCTACCGCA GCCGCATCGA CNACTACATC CACAACGTTT ACGGGAAATG
               1801
                     GTGGGATTTG AACGGGAATA TTCCGAGCTG GGTCAGCAGC ACCGGGCTTG
               1851
 55
                     CCTACACCAT CCAACACCGC AATTTCAAAG ACAAAGTGCA CAAACACGGT
               1901
                     TTTGAGTTGG AGCTGAATTA CGATTATNGG CGTTTTTCA CCAACCTTTC TTACGCCTAT CAAAAAAGCA CGCAACCGAC CAACTTCAGC GATGCGAGCG
               1951
               2001
               2051
                     AATCGCCCAA CAATGCGTCC AAAGAAGACC AACTCAAACA AGGTTATGGG
                     TTGAGCAGGG TTTCCGCCCT GCCGCGAGAT TACGGACGTT TGGAAGTCGG
               2101
 60
                     TACGCGCTGG TTGGGCAACA AACTGACTTT GGGCGGCGCG ATGCGCTATT
               2151
                2201
                     TCGGCAAGAG CATCCGCGCG ACGGCTGAAG AACGCTATAT CGACGNCACC
                     AATGGGGNAN NTACCAGCAA TTTCCGGCAA CTGGGCAAGC GTTCCATCAN
               2251
                2301
                     ACAAACCGAA ACCCTTGCCC GCCAGCCTTT GATTTTTGAT TTNTACGCCG
                     CTTACGAGCC GAAGAAAAAN CTTATTTTCC GCGCCGAAGT CAAAAATCTG
                2351
 65
                     TTCGACAGGC GTTATATCGA TCCGCTCGAT GCGGGCAATG ATGCGGCAAC
                2401
                     GCAGCGTTAT TACAGTTCGT TCGACCCGAA AGACAAGGAC GAAGAAGTAA
                2451
                     CGTGTAATGA TGATAACACG TTATGCAACG GCAAATACGG CGGCACAAGC
                2501
                     AAAAGCGTAT TGACCAATTT TGCACGCGGA CNCACCTTTT TGATAACGAT
                2551
```

GAGCTACAAG TTTTAA

2601

This encodes a protein having (partial) amino acid sequence <SEQ ID 880>:

	Tims cheodes a protein	
5 10 15	51 RXDS 101 VKGS 151 NAMA 201 ERRK 251 EGHD 301 DLNT 351 GWGI 401 FPEE 451 YFDA 501 KHCN 551 NIQE 601 GYRS 651 FELE 701 LSRV 751 NGX	CVFTDAR AVSTRODIFK SXENLDNIVR XIPGAFTXQX KSSGXVSLNI GFGRVN TMVDGITXTF YSTSTDAGRA GGSSQFGASV DSNFXAGLDV GFSGSAG INSLAGSANL RTLXVDDVVQ GNXTYGLLLK GLTGTNSTKG LAIGARK WLESGASVGV LYGHSRRSVA QNYRVGGGGQ HIGNFGAEYL CORYFEQ EGGLKFNSNS GKWERDFQKS YWKTKWYQKY DAPQELQKYI DFSSLKXQSA GNLFKLEYDG VFNKYTAQFR CKIGSRK IINRNYQFNY GLSLNPYTNL NLTAAYNSGR QKYPKGSKFT YNNAKILDLX NTSTFRLPRE TELQTTLGFN YFHNEYGKNR CLGLFFD GPDXDNGLYS YLGRFKGDKG LLPQKSTIVQ PAGSQYFNTF AALKKDI YRLNYSTNTV GYRFGGXYTG YYXSDDEFKR AFGENSPTYX NQSCGIY EPVLKKYGKK RANNHSVSIS ADFGDYFMPF ASYSRTHRMP CMYFSQI GDSGVHTALK PERANTWQFG FNTYKKGLLK QDDILGLKLV GRIDXYI HNVYGKWWDL NGNIPSWVSS TGLAYTIQHR NFKDKVHKHG CLNYDYX RFFTNLSYAY QKSTQPTNFS DASESPNNAS KEDQLKQGYG VSALPRD YGRLEVGTRW LGNKLTLGGA MRYFGKSIRA TAEERYIDXT KTSNFRQ LGKRSIXQTE TLARQPLIFD XYAAYEPKKX LIFRAEVKNL
		RYIDPLD AGNDAATORY YSSFDPKDKD EEVTCNDDNT LCNGKYGGTS LTNFARG XTFLITMSYK F*
	• • • • • • • • • • • • • • • • • • • •	
20	ORF133a and ORF1	33-1 show 94.3% identity in 871 aa overlap:
		10 20 30 40
	orf133a.pep	KDKKVFTDARAVSTRQDIFKSXENLDNIVRXIPGAFTXQXKS
	<u> </u>	
25	orf133-1	EAQIQVLEDVHVKAKRVPKDKKVFTDARAVSTRQDIFKSSENLDNIVRSIPGAFTQQDKS 10 20 30 40 50 60
23		
		50 60 70 80 90 100
	orf133a.pep	SGXVSLNIRXDSGFGRVNTMVDGITXTFYSTSTDAGRAGGSSQFGASVDSNFXAGLDVVK
30	orf133-1	SGIVSLNIRGDSGFGRVNTMVDGITQTFYSTSTDAGRAGGSSQFGASVDSNFIAGLDVVK
		70 80 90 100 110 120
35	orf133a.pep	110 120 130 140 150 160 GSFSGSAGINSLAGSANLRTLXVDDVVQGNXTYGLLLKGLTGTNSTKGNAMAAIGARKWL
		130 140 150 160 170 180
40	orf133a.pep	170 180 190 200 210 220 ESGASVGVLYGHSRRSVAQNYRVGGGGQHIGNFGAEYLERRKQRYFEQEGGLKFNSNSGK
	orf133-1	ESGASVGVLYGHSRRSVAQNYRVGGGGQHIGNFGAEYLERRKQRYFVQEGALKFNSDSGK 190 200 210 220 230 240
45	orf133a.pep orf133-1	230 240 250 260 270 280 WERDFQKSYWKTKWYQKYDAPQELQKYIEGHDKSWRENLAPQYDITPIDPSSLKXQSAGN
50		250 260 270 280 290
50		290 300 310 320 330 340
	orf133a.pep	LFKLEYDGVFNKYTAQFRDLNTKIGSRKIINRNYQFNYGLSLNPYTNLNLTAAYNSGRQK
	orf133-1	
55	011133-1	300 310 320 330 340 350
		200 200 200 200 400
	orf133a.pep	350 360 370 380 390 400 YPKGSKFTGWGLXKDFETYNNAKILDLXNTSTFRLPRETELQTTLGFNYFHNEYGKNRFP
60	orf133-1	YPKGSKFTGWGLLKDFETYNNAKILDLNNTATFRLPRETELQTTLGFNYFHNEYGKNRFP 360 370 380 390 400 410
		200 270 200 320 320
		410 420 430 440 450 460
65	orf133a.pep	EELGLFFDGPDXDNGLYSYLGRFKGDKGLLPQKSTIVQPAGSQYFNTFYFDAALKKDIYR
03	orf133-1	EELGLFFDGPDQDNGLYSYLGRFKGDKGLLPQKSTIVQPAGSQYFNTFYFDAALKKDIYR
		420 430 440 450 460 470

-478-

	orf133a.pep		170 JTVGVDEGG	480	490	500 PTYXKHCNOS	510	520
	Offissa. beb		111111111			:		
5	orf133-1	LNYSTN 480	NTVGYRFGG 490	EYTGYYGSDI 500		PTYKKHCNRS		KYGKKRA
3		480	490	500	510	520	530	
		-	30	540	550	560	570	580
	orf133a.pep					FSQIGDSGVH!!!!		
10	orf133-1	NNHSV	SISADFGDY	(FMPFASYSRI	HRMPNIQEM	FSQIGDSGVH		
		540	550	560	570	580	590	
		!	590	600	610	620	630	640
16	orf133a.pep					WWDLNGNIPS		
15	orf133-1					:    WWDLNGDIPS		
	022200	600	610	620	630	640	650	114
			650	660	670	680	690	700
20	orf133a.pep	KDKVHI	KHGFELELN	NYDYXRFFTNI	SYAYQKSTQ	TNFSDASESE	NNASKEDOL	KOGYGLS
	£122 1	11111	MCEELELN	IIII IIIII				1111111
	orf133-1	660	670	680	-51A1QK5TQI 690	TNFSDASESE 700	PNNASKEDQL 710	KQGYGLS
25								
25	orf133a.pep		710 PRDYGRLEV	720 ZGTRWLGNKL	730 LIGGAMRYFGI	740 KSIRATAEERY	750 (TDXTNGXXT	760 SNEROLG
•	•	11111			пинин		111 111 1	шіш
	orf133-1	RVSAL	PRDYGRLEV 730	VGTRWLGNKL: 740	rlggamryfgi 750	SIRATAEERY 760	IDGTNGGNT 770	SNFRQLG
30	•				750	700	770	
	orf133a.pep		770	780	790	800 EVKNLFDRRYI	810	820
	Offissa.pep							
25	orf133-1				EPKKNLIFRA	EVKNLFDRRY	IDPLDAGNDA	
35		780	790	800	810	820	830	
			830	840	850	860	· 870	
	orf133a.pep					NFARGXTFLIT		
40	orf133-1	SFDPK	DKDEDVTC	NADKTLCNGK	YGGTSKSVLT	NFARGRTFLM		
		840	850	860	870	880		÷

# Homology with a predicted ORF from N.gonorrhoeae

ORF133 shows 92.3% identity over 392 aa overlap with a predicted ORF (ORF133ng) from N.

# 45 gonorrhoeae:

	orf133.pep	PGYYGSDDEFKRAFGENSPTXKKHCNRSCGI	31
	orf133ng	FYFDAALKKDIYRLNYSTNAINYRFGGEYTGYYGSENEFKRAFGENSPAYKEHCDPSCGL	560
50	orf133.pep	YEPVLKKYGKKRANNHSVSISADFGDYFMPFASYSRTHRMPNIQEMYFSQIGDSGVHTAL	91
	orf133ng	YEPVLKKYGKKRANNHSVSISADFGDYFMPFAGYSRTHRMPNIQEMYFSQIGDSGVHTAL	620
55	orf133.pep	KPERANTWQFGFXTYKKGLLKQDDTLGLKLVGYRSRIDNYIHNVYGKWWDLNGDIPSWVS	151
	orf133ng	KPERANTWQFGFNTYKKGLLKQDDILGLKLVGYRSRIDNYIHNVYGKWWDLNGDIPSWVG	680
	orf133.pep	STGLAYTIQHRXFXDKVHQXXXXXXXYDYGRFFTNLSYAYQKSTQPTNFSDASESPNNA	211
60	orf133ng	STGLAYTIRHRNFKDKVHKHGFELELNYDYGRFFTNLSYAYQKSTQPTNFSDASESPNNA	740
	orf133.pep	SKEDQLKQGYGLSRVSALPRDYGRLEVGTRWLGNKLTLGGAMRYFGKSIRATAEERYIDG	271
65	orf133ng	SKEDQLKQGYGLSRVSALPRDYGRLEVGTRWLGNKLTLGGAMRYFGKSIRATAEERYIDG	800
	orf133.pep	TNGGNTSNFRQLGKRSIKQTETLARQPLIXDFNAAYEPKKNLIFRAEVKNLFDRRYIDPL	331
	orf133ng	TNGGNTSNVRQLGKRSIKQTETLARQPLIFDFYAAYEPKKNLIFRAEVKNLFDRRYIDPL	860

The complete length ORF133ng nucleotide sequence <SEQ ID 881> is predicted to encode a protein having amino acid sequence <SEQ ID 882>:

```
MRSSFRLKPI CFYLMGVMLY HHSYAEDAGR AGSEAQIQVL EDVHVKAKRV
                       PKDKKVFTDA RAVSTRQDVF KSGENLDNIV RSIPGAFTQQ DKSSGIVSLN
                       IRGDSGFGRV NTMVDGITQT FYSTSTDAGR AGGSSQFGAS VDSNFIAGLD
                       VVKGSFSGSA GINSLAGSAN LRTLGVDDVV QGNNTYGLLL KGLTGTNSTK
GNAMAAIGAR KWLESGASVG VLYGHSRRGV AQNYRVGGGG QHIGNFGEEY
15
                       LERRKOOYFV QEGGLKFNAG SGKWERDLQR QYWKTKWYKK YEDPQELQKY
                 251
                       IEEHDKSWRE NLAPQYDITP IDPSGLKQQS AGNLLNLEYD GVFNKYTAQF
                       RDLNTRIGSR KIINRNYQFN YGLSLNPYTN LNLTAAYNSG RQKYPKGAKF
                  351
                       TGWGLLKDFE TYNNAKILDL NNTATFRLPR ETELQTTLGF NYFHNEYGKN
                  401
                  451
                       RFPEELGLFF DGPDQDNGLY SYLGRFKGDK GLLPQKSTIV QPAGSQYFNT
20
                       FYFDAALKKD IYRLNYSTNA INYRFGGEYT GYYGSENEFK RAFGENSPAY
                  501
                       KEHCDPSCGL YEPVLKKYGK KRANNHSVSI SADFGDYFMP FAGYSRTHRM
                  551
                       PNIQEMYFSQ IGDSGVHTAL KPERANTWQF GFNTYKKGLL KQDDILGLKL
                  601
                       VGYRSRIDNY IHNVYGKWWD LNGDIPSWVG STGLAYTIRH RNFKDKVHKH
                  651
                       GFELELNYDY GRFFTNLSYA YQKSTQPTNF SDASESPNNA SKEDQLKQGY
25
                       GLSRVSALPR DYGRLEVGTR WLGNKLTLGG AMRYFGKSIR ATAEERYIDG
TNGGNTSNVR QLGKRSIKQT ETLARQPLIF DFYAAYEPKK NLIFRAEVKN
                  751
                  801
                       LFDRRYIDPL DAGNDAATQR YYSSFDPKDK DEDVTCNADK TLCNGKYGGT
                  851
                       SKSVLTNFAR GRTFLMTMSY KF*
```

30 A variant was also identified, being encoded by the gonococcal DNA sequence <SEQ ID 883>:

```
ATGAGATCTT CTTTCCGGTT GAAGCCGATT TGTTTTTATC TTATGGGTGT
                    TATGCTATAT CATCATAGTT ATGCCGAAGA TGCAGGGCGC GCGGCCAGCG
                    AGGCGCAGAT ACAGGTTTTG GAAGATGTGC ACGTCAAGGC GAAGCGCGTA
               101
                    CCGAAAGACA AAAAAGTGTT TACCGATGCG CGTGCCGTAT CGACCCGTca
                    ggatgtgttc aaatccggcg aaaacctcga caacatcgta cgcagcatac
35
               201
                    CCGGTGCGTT TACACAGCAA GATAAAAGCT CGGGCATTGT GTCTTTGAAT
                    ATTCGCGGCG ACAGCGGGTT CGGGCGGGTC AATACGATGG TGGACGGCAT
                301
                    CACGCAGACC TTTTATTCGA CTTCTACCGA TGCGGGCAGG GCAGGCGGTT
                    CATCTCAATT CGGTGCATCT GTCGACAGCA ATTTTATTGC CGGACTGGAT
                401
                    GTCGTCAAAG GCAGCTTCAG CGGCTCGGCA GGCATCAACA GCCTTGCCGG
40
                451
                    TTCGGCGAAT CTGCGGACTT TAGGCGTGGA TGACGTCGTT CAGGGCAATA
                501
                    ATACCTACGG CCTGCTGCTA AAAGGTCTGA CCGGCACCAA TTCAACCAAA
                551
                601
                    GGTAATGCGA TGGCGGCGAT AGGTGCGCGC AAATGGCTGG AAAGCGGAGC
                    GTCTGTCGGT GTGCTTTACG GGCACAGCAG GCGCGGCGTG GCGCAAAATT
                    ACCGCGTGGG CGGCGGGG CAGCACATCG GAAATTTTGG TGAAGAATAT
45
                701
                     CTGGAACGGC GCAAACAGCA ATATTTTGTA CAAGAGGGTG GTTTGAAATT
                801 CAATGCCGGC AGCGGAAAAT GGGAACGGGA TTTGCAAAGG CAATACTGGA
                851 AAACAAAGTG GTATAAAAAA TACGAAGACC CCCAAGAACT GCAAAAATAC
                    ATCGAAGAGC ATGATAAAAG CTGGCGGGAA AACCTGGCGC CGCAATACGA
                901
                     CATCACCCCC ATCGATCCGT CCGGCCTGAA GCAGCAGTCG GCAGGCAATC
50
                951
                     TGTTTAAATT GGAATACGAC GGCGTATTCA ATAAATACAC GGCGCAATTT
               1001
                     CGCGATTTAA ACACCAGAAT CGGCAGCCGC AAAATCATCA ACCGCAATTA
               1051
                     TCAATTCAAT TACGGTTTGT CTTTGAACCC GTATACCAAC CTCAATCTGA
               1101
                     CCGCAGCCTA CAATTCGGGC AGGCAGAAAT ATCCGAAAGG GGCGAAGTTT
               1151
                     ACAGGCTGGG GGCTTTTAAA AGATTTTGAA ACCTACAACA ACGCGAAAAT
55
               1201
                     CCTCGACCTC AACAACACCG CCACCTTCCG GCTGCCCCGC GAAACCGAGT
               1251
                     TGCAAACCAC TTTGGGCTTC AATTATTTCC ACAACGAATA CGGCAAAAAC
               1301
                     CGCTTTCCTG AAGAATTGGG GCTGTTTTTC GACGGTCCTG ATCAGGACAA
               1351
                     CGGGCTTTAT TCCTATTTGG GGCGGTTTAA GGGCGATAAA GGGCTGTTGC
               1401
                     CTCAAAAATC AACCATTGTC CAACCGGCCG GCAGCCAATA TTTCAACACG
60
               1451
                     TTCTACTTCG ATGCCGCGCT CAAAAAAGAC ATTTACCGCT TAAACTACAG
               1501
                     CACCAATGCA ATCAACTACC GTTTCGGCGG CGAATATACG GGCTATTACG
               1551
                     GCTCGGAAAA CGAATTTAAG CGGGCATTCG GAGAAAACTC GCCGGCATAC
               1601
               1651 AAGGAACATT GCGACCCGAG CTGCGGGCTT TATGAACCCG TATTGAAAAA
               1701 ATACGGCAAA AAGCGCGCCA ACAACCATTC GGTCAGCATT AGTGCGGACT
 65
               1751 TCGGCGATTA TTTCATGCCG TTCGCCGGCT ATTCGCGCAC ACACCGTATG
```

	1801 CCCAACATCC AAGAAATGTA TTTTTCCCAA ATCGGCGACT CCGGCGTTCA
	1851 CACCGCCTTA AAACCAGAGC GCGCAAACAC TTGGCAATTT GGCTTCAATA
	1901 CCTATAAAAA AGGATTGTTA AAACAAGATG ATATATTAGG ATTGAAACTG
5	1951 GTCGGCTACC GCAGCCGCAT TGACAACTAC ATCCACAACG TTTACGGGAA
3	2001 ATGGTGGGAT TTGAACGGGG ATATTCCGAG CTGGGTCGGC AGCACCGGGC 2051 TTGCCTACAC CATCCGACAC CGCAATTTCA AAGACAAAGT GCACAAACAC
	2001 INGCCIACAC CANCEGRACAC COCAAANTICA AAGACAAAGI GCACAAACAC 2101 GGTTTTGAGC TGGAGCTGAA TTACGATTAT GGGCGTTTTT TCACCAACCT
	2151 TTCTTACGCC TATCAAAAAA GCACGCAACC GACCAATTTC AGCGATGCGA
	2201 GCGAATCGCC CAACAATGCC tccaaAGAAG ACCAACTCAA ACAAGGTTAT
10	2251 GGGCTGAGCA GGGTTTCCGC CCTGCCGCGA GATTACGGAC GTTTGGAAGT
	2301 CGGTACGCGC TGGTTGGGCA ACAAACTGAC TTTGGGCGGC GCGAtgcGCT
	2351 ATTTCGGCAA GAGCATCCGC GCGACGGCTG AAGAACGCTA TATCGACGGC 2401 ACCAACGGGG GAAATACCAG CAATGTCCGG CAACTGGGCA AGCGTTCCAT
	2401 ACCAACGGGG GAAATACCAG CAATGTCCGG CAACTGGGCA AGCGTTCCAT 2451 CAAACAAACC GAAACCCTTG CCCGACAGCC TTTGATTTTT GATTTTTACG
15	2501 CCGCTTACGA GCCGAAGAAA AACCTTATTT TCCGCGCCGA AGTCAAAAAC
	2551 CTGTTCGACA GGCGTTATAT CGATCCGCTC GATGCGGGCA ATGATGCGGC
	2601 AACGCAGCGT TATTACAGCT CGTTCGACCC GAAAGACAAG GACGAAGACG
	2651 TAACGTGTAA TGCTGATAAA ACGTTGTGCA ACGGCAAATA CGGCGGCACA
20	2701 AGCAAAAGCG TATTGACCAA TTTCGCACGC GGACGCACCT TCTTGATGAC
20	2751 GATGAGCTAC AAGTTTTAA
	This corresponds to the amino acid sequence <seq 884;="" id="" orf133ng-1="">:</seq>
	1 MRSSFRLKPI CFYLMGVMLY HHSYAEDAGR AGSEAQIQVL EDVHVKAKRV
	51 PKDKKVFTDA RAVSTRODVF KSGENLDNIV RSIPGAFTQQ DKSSGIVSLN
25	101 IRGDSGFGRV NTMVDGITQT FYSTSTDAGR AGGSSQFGAS VDSNFIAGLD
25	151 VVKGSFSGSA GINSLAGSAN LRTLGVDDVV QGNNTYGLLL KGLTGTNSTK
	201 GNAMAAIGAR KWLESGASVG VLYGHSRRGV AQNYRVGGGG QHIGNFGEEY 251 LERRKQQYFV QEGGLKFNAG SGKWERDLOR QYWKTKWYKK YEDPOELOKY
	301 IEEHDKSWRE NLAPQYDITP IDPSGLKQQS AGNLFKLEYD GVFNKYTAOF
	351 RDLNTRIGSR KIINRNYQFN YGLSLNPYTN LNLTAAYNSG RQKYPKGAKF
30	401 TGWGLLKDFE TYNNAKILDL NNTATFRLPR ETELQTTLGF NYFHNEYGKN
	451 RFPEELGLFF DGPDQDNGLY SYLGRFKGDK GLLPQKSTIV QPAGSQYFNT
	501 FYFDAALKKD IYRLNYSTNA INYRFGGEYT GYYGSENEFK RAFGENSPAY 551 KEHCDPSCGL YEPVLKKYGK KRANNHSVSI SADFGDYFMP FAGYSRTHRM
•	601 PNIQEMYFSQ IGDSGVHTAL KPERANTWOF GFNTYKKGLL KODDILGLKL
35	651 VGYRSRIDNY IHNVYGKWWD LNGDIPSWVG STGLAYTIRH RNFKDKVHKH
	701 GFELELNYDY GRFFTNLSYA YQKSTQPTNF SDASESPNNA SKEDQLKQGY
	751 GLSRVSALPR DYGRLEVGTR WLGNKLTLGG AMRYFGKSIR ATAEERYIDG
	801 TNGGNTSNVR QLGKRSIKQT ETLARQPLIF DFYAAYEPKK NLIFRAEVKN
40	851 LFDRRYIDPL DAGNDAATOR YYSSFDPKDK DEDVTCNADK TLCNGKYGGT 901 SKSVLTNFAR GRTFLMTMSY KF*
40	
	ORF133ng-1 and ORF133-1 show 96.2% identity in 889 aa overlap:
	10 20 30 40 50 60
•	orf133ng-1.pep SFRLKPICFYLMGVMLYHHSYAEDAGRAGSEAQIQVLEDVHVKAKRVPKDKKVFTDARAV
45	orf133-1
45	10 20 30
	70 80 90 100 110 120
50	orf133ng-1.pep STRQDVFKSGENLDNIVRSIPGAFTQQDKSSGIVSLNIRGDSGFGRVNTMVDGITQTFYS
50	orf133-1 STRQDIFKSSENLDNIVRSIPGAFTQQDKSSGIVSLNIRGDSGFGRVNTMVDGITOTFYS
	40 50 60 70 80 90
55	130 140 150 160 170 180 orf133ng-1.pep TSTDAGRAGGSSQFGASVDSNFIAGLDVVKGSFSGSAGINSLAGSANLRTLGVDDVVQGN
55	0::1:35mg-1.pep 131bAGAAGG35QEGA3VBME1AGEDVVAGSESGSAGIMSEAGSAMEKIEGVDDVVQGM
	orf133-1 TSTDAGRAGGSSQFGASVDSNFIAGLDVVKGSFSGSAGINSLAGSANLRTLGVDDVVQGN
	100 110 120 130 140 150
60	100 000 000
OU	190 200 210 220 230 240
	orf133ng-1.pep NTYGLLLKGLTGTNSTKGNAMAAIGARKWLESGASVGVLYGHSRRGVAQNYRVGGGGQHI
	orf133-1 NTYGLLLKGLTGTNSTKGNAMAAIGARKWLESGASVGVLYGHSRRSVAQNYRVGGGGQHI
	160 170 180 190 200 210
65	250 260 000 000
	250 260 270 280 290 300 orf133ng-1.pep GNFGEEYLERRKQQYFVQEGGLKFNAGSGKWERDLQRQYWKTKWYKKYEDPQELQKYIEE
	ATTIONS TIPED GUIGHDINGWARITAGOGUURWADGUARUNDAUALULULIULIULIULIULULIULULIULULULULULUL

	orf133-1	 GNFGAEYLERF 220	:     :   RKQRYFVQEGALK 230	:	 DLQRQQWKYKI 250	: ::     PYKNYNN-QE   260	LOKYIEE
5	orf133ng-1.pep	11111111	320 PQYDITPIDPSGI            PQYDITPIDPSSI	 KQQSAGNLFKI	EYDGVFNKY	  TAQFRDLNTK	111111
10		270 280 370 NRNYOFNYGL	380	300 390 YNSGROKYPKO	310 400 SAKFTGWGLL	320 410 KDFETYNNAK	420 ILDLNNT
15	orf133-1			111111111	:	KDFETYNNAK 380	ILDLNNT
20	orf133ng-1.pep	11111111111			1111111111	111111111	11111
25	orf133ng-1.pe	111111111		5   5   5   6   7   7   7   7   7   7   7   7   7	111:::1111	11111111	::!!!!!!
30	orf133ng-1.pe	11111:11:1	:    :       CNRSCGIYEPVL	11111111111	11111111		1111111
35	orf133ng-1.pe	 QEMYFSQIGI			 KGLLKQDDTI	LGLKLVGYRS	111111
40	orf133ng-1.pe	570 58 670 p VYGKWWDLNO	680	600 690 YTIRHRNFKDF	610 700 WHKHGFELEI	620 710 LNYDYGRFFT	720 NLSYAYQK
45	orf133-1	VYGKWWDLNG	GDIPSWVSSTGLA 40 650 740	YTIQHRNFKDI 660 750	(VHKHGFELE) 670 760	LNYDYGRFFT 680 770	NLSYAYQK 780
50	orf133ng-1.pe	STQPTNFSDA          STQPTNFSDA	ASESPNNASKEDO 	LKQGYGLSRV	111111	11111111	111111
55	orf133ng-1.pe	 YFGKSIRAT	800 AEERYIDGTNGGN            AEERYIDGTNGGN 60 770			111111111	41111111
60	orf133ng-1.pd	 FRAEVKNLF	860 DRRYIDPLDAGNI            DRRYIDPLDAGNI 20 830	1111111111	111111111	11111111	
65	orf133ng-1.p		920 TFLMTMSYKFX           TFLMTMSYKFX  80				

70 In addition, ORF133ng-1 is homologous to a TonB-dependent receptor in *H.influenzae*:

5	<pre>&gt;gi 107537 Haemophilu protein 1 Score =</pre>	YC17_HAEIN PROBABLE TONB-DEPENDENT RECEPTOR HI1217 PRECURSOR $2 pir  664110$ transferrin binding protein 1 precursor (tbp1) has influenzae (strain Rd KW20) >gi 1574147 (U32801) transferrin precursor (tbp1) [Haemophilus influenzae] Length = 913 930 bits (2377), Expect = 0.0 s = 476/921 (51%), Positives = 619/921 (66%), Gaps = 72/921 (79)	binding
4.0		QVLEDVHVKAKRVPKDKKVFTDARAVSTRQDVFKSGENLDNIVRSIPGAFTQQDKSSGIV + L + V K + DKK FT+A+A STR++VFK + +D ++RSIPGAFTQQDK SG+V	
10	Sbjct: 29	ETLGQIDVVEKVISNDKKPFTEAKAKSTRENVFKETQTIDQVIRSIPGAFTQQDKGSGVV	88
	Query: 98	SLNIRGDSGFGRVNTMVDGITQTFYSTSTDAGRAGGSSQFGASVDSNFIAGLDVVKGSFS S+NIRG++G GRVNTMVDG+TQTFYST+ D+G++GGSSQFGA++D NFIAG+DV K +FS	157
15	Sbjct: 89		148,
	Query: 158	GSAGINSLAGSANLRTLGVDDVVQXXXXXXXXXXXXXXXXXXXXXAMAAIGARKWLESGA G++GIN+LAGSAN RTLGV+DV+ M RKWL++G	217
	Sbjct: 149	GASGINALAGSANFRTLGVNDVITDDKPFGIILKGMTGSNATKSNFMTMAAGRKWLDNGG	208
20	Query: 218	SVGVLYGHSRRGVAQNYRVGGGGQHIGNFGEEYLERRKQQYFVQEGGLKFNAGSGKWERD VGV+YG+S+R V+Q+YR+ GGG+ + + G++ L + K+ YF + G N G+W D	277
	Sbjct: 209	YVGVVYGYSQREVSQDYRI-GGGERLASLGQDILAKEKEAYF-RNAGYILNP-EGQWTPD	265 <sup>-</sup>
25	Query: 278	LQRQYWKTKWYKKYEDPQELQKYIEE	303
	Sbjct: 266	L +++W +Y KK +D ++LQK IEE LSKKHWSCNKPDYQKNGDCSYYRIGSAAKTREILQELLTNGKKPKDIEKLQKGNDGIEE	325
	Query: 304	HDKSWRENLAPQYDITPIDPSGLKQQSAGNLFKLEYDGVFNKYTAQFRDLNTRIGSRKII	363
30	Sbjct: 326	DKS+ N QY + PI+P L+ +S +L K EY AQ R L+ +IGSRKI TDKSFERN-KDQYSVAPIEPGSLQSRSRSHLLKFEYGDDHQNLGAQLRTLDNKIGSRKIE	384
	Query: 364	NRNYQFNYGLSLNPYTNLNLTAAYNSGRQKYPKGAKFTGWGLLKDFETYNNAKILDLNNT	423
35	Sbjct: 385	NRNYQ NY + N Y +LNL AA+N G+ YPKG F GW + T N A I+D+NN+ NRNYQVNYNFNNNSYLDLNLMAAHNIGKTIYPKGGFFAGWQVADKLITKNVANIVDINNS	444
33	Query: 424	ATFRLPRETELQTTLGFNYFHNEYGKNRFPEELGLFFDGPDQDNGLYSYLGRFKGDKG	481
	Sbjct: 445	TF LP+E +L+TTLGFNYF NEY KNRFPEEL LF++ D GLYS+ GR+ G K HTFLLPKEIDLKTTLGFNYFTNEYSKNRFPEELSLFYNDASHDQGLYSHSKRGRYSGTKS	504
40	Query: 482	LLPQKSTIVQPAGSQYFNTFYFDAALKKDIYRLNYSTNAINYRFGGEYTGYYGSENEFKR LLPQ+S I+QP+G Q F T YFD AL K IY LNYS N +Y F GEY GY	541
	Sbjct: 505	LLPQRSVILQPSGKQKFKTVYFDTALSKGIYHLNYSVNFTHYAFNGEYVGY	555
45	Query: 542	2 AFGENSPAYKEHCDPSCGLYEPVLKKYGKKRANNHSVSISADFGDYFMPFAGYSRTHRMP EN+ + + EP+L K G K+A NHS ++SA+ DYFMPF YSRTHRMP	601
40	Sbjct: 556	EN+ + + EP+L K G K+A NHS ++SA+ DYFMPF YSRTHRMP 5ENTAGQQINEPILHKSGHKKAFNHSATLSAELSDYFMPFFTYSRTHRMP	604
	Query: 602	NIQEMYFSQIGDSGVHTALKPERANTWQFGFNTYKKGLLKQDDILGLKLVGYRSRIDNYI	661
50	Sbjct: 605	NIQEM+FSQ+ ++GV+TALKPE+++T+Q GFNTYKKGL QDD+LG+KLVGYRS I NYI 5 NIQEMFFSQVSNAGVNTALKPEQSDTYQLGFNTYKKGLFTQDDVLGVKLVGYRSFIKNYI	664
	Query: 662	HNVYGKWWDLNGDIPSWVGSTGLAYTIRHRNFKDKVHKHGFELELNYDYGRFFTNLSYAY HNVYG WW +P+W S G YTI H+N+K V K G ELE+NYD GRFF N+SYAY	721
55	Sbjct: 669	HNVYG WW +P+W S G YTI H+N+K V K G ELE+NYD GRFF N+SYAY 5 HNVYGVWWRDGMPTWAESNGFKYTIAHQNYKPIVKKSGVELEINYDMGRFFANVSYAY	722
33	Query: 722	QKSTQPTNFSDASESPNNASKEDQLKQGYGLSRVSALPRDYGRLEVGTRWLGNKLTLGGA	781
	Sbjct: 723	Q++ QPTN++DAS PNNAS+ED LKQGYGLSRVS LP+DYGRLE+GTRW KLTLG A QRTNQPTNYADASPRPNNASQEDILKQGYGLSRVSMLPKDYGRLELGTRWFDQKLTLGLA	782
60	Query: 782	MRYFGKSIRATAEERYIDGTNGGNTSNVRQLGKRSIKQTETLARQPLIFDFYAAYEPKKN	841
	Sbjct: 78	RY+GKS RAT EE YI+G+ + +R+ ++K+TE + +QP+I D + +YEP K+ 3 ARYYGKSKRATIEEEYINGSR-FKKNTLRRENYYAVKKTEDIKKQPIILDLHVSYEPIKD	841
65	Query: 842	2 LIFRAEVKNLFDRRYIDPLDAGNDAATQRYYSSFDPKDKDEDVTCNADKTLCNGKYGGTS	901
65	Sbjct: 842	LI +AEV+NL D+RY+DPLDAGNDAA+QRYYSS + + C D + C GG+ LIIKAEVQNLLDKRYVDPLDAGNDAASQRYYSSLNNSIECAQDSSACGGSD	892
	Query: 90	2 KSVLTNFARGRTFLMTMSYKF 922	
70	Sbjct: 89	K+VL NFARGRT++++++YKF 3 KTVLYNFARGRTYILSLNYKF 913	
		·	

The underlined motif in the gonococcal protein (also present in the meningococcal protein) is predicted to be an ATP/GTP-binding site motif A (P-loop), and the analysis suggests that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

## 5 Example 104

The following partial DNA sequence was identified in N.meningitidis <SEQ ID 885>

```
ATGAACCTGA TTTCACGTTA CATCATCCGT CAAATGGCGG TTATGGCGGT
                     TTACGCGCTC CTTGCCTTCC TCGCTTTGTA CAGCTTTTTT GAAATCCTGT
                     ACGAAACCGG CAACCTCGGC AAAGGCAGTT ACGGCATATG GGAAATGCTG
               101
                     GGCTACACCG CCCTCAAAAT GCCCGCCCGC GCCTACGAAC TGATTCCCCT
10
                     CGCCGTCCTT ATCGGCGGAC TGGTCTCCCT CAGCCAGCTT GCCGCCGGCA
                201
                     GCGAACTGAC CGTCATCAAA GCCAGCGGCA TGAGCACCAA AAAGCTGCTG
                251
                     TTGATTCTGT CGCAGTTCGG TTTTATTTTT GCTATTGCCA CCGTCGCGCT
                301
                     CGGCGAATGG GTTGCGCCCA CACTGAGCCA AAAAGCCGAA AACATCAAAG
                351
                     CCGCCGCCAT CAACGGCAAA ATCAGCACCG GCAATACCGG CCTTTGGCTG
15
                401
                     AAAGAAAAA ACAGCGTGAT CAATGTGCGC GAAATGTTGC CCGACCAT..
                451
```

This corresponds to the amino acid sequence <SEQ ID 886; ORF112>:

```
MNLISRYIIR QMAVMAVYAL LAFLALYSFF EILYETGNLG KGSYGIWEML
51 GYTALKMPAR AYELIPLAVL IGGLVSLSQL AAGSELTVIK ASGMSTKKLL
20 101 LILSQFGFIF AIATVALGEW VAPTLSQKAE NIKAAAINGK ISTGNTGLWL
151 KEKNSVINVR EMLPDH...
```

Further work revealed further partal nucleotide sequence <SEQ ID 887>:

```
ATGAACCTGA TTTCACGTTA CATCATCCGT CAAATGGCGG TTATGGCGGT
                     TTACGCGCTC CTTGCCTTCC TCGCTTTGTA CAGCTTTTTT GAAATCCTGT
                 51
                     ACGAAACCGG CAACCTCGGC AAAGGCAGTT ACGGCATATG GGAAATGCTG
25
                101
                     gGCTACACCG CCCTCAAAAT GCCCGCCCGC GCCTACGAAC TGATTCCCCT
                     CGCCGTCCTT ATCGGCGGAC TGGTCTCCCT CAGCCAGCTT GCCGCCGGCA
                201
                     GCGAACTGAC CGTCATCAAA GCCAGCGGCA TGAGCACCAA AAAGCTGCTG
                251
                     TTGATTCTGT CGCAGTTCGG TTTTATTTTT GCTATTGCCA CCGTCGCGCT
                     CGGCGAATGG GTTGCGCCCA CACTGAGCCA AAAAGCCGAA AACATCAAAG
30
                351
                     CCGCCGCCAT CAACGGCAAA ATCAGCACCG GCAATACCGG CCTTTGGCTG
                401
                     AAAGAAAAA ACAGCTTKAT CAATGTGCGC GAAATGTTGC CCGACCATAC
                451
                     GCTTTTGGGC ATCAAAATTT GGGCGCGCAA CGATAAAAAC GAATTGGCAG
                501
                     AGGCAGTGGA AGCCGATTCC GCCGTTTTGA ACAGCGACGG CAGTTGGCAG
                551
                     TTGAAAAACA TCCGCCGCAG CACGCTTGGC GAAGACAAAG TCGAGGTCTC
35
                601
                     TATTGCGGCT GAAGAAAACT GGCCGATTTC CGTCAAACGC AACCTGATGG
                651
                     ACGTATTGCT CGTCAAACCC GACCAAATGT CCGTCGGCGA ACTGACCACC
                     TACATCCGCC ACCTCCAAAA CAACAGCCAA AACACCCGAA TCTACGCCAT
                751
                     CGCATGGTGG CGCAAATTGG TTTACCCCGC CGCAGCCTGG GTGATGGCGC
                     TCGTCGCCTT TGCCTTTACC CCGCAAACCA CCCGCCACGG CAATATGGGC
40
                851
                     TTAAAACTCT TCGGCGGCAT CTGTsTCGGA TTGCTGTTCC ACCTTGCCGG
                901
                951 ACGGCTCTTT GGGTTTACCA GCCAACTCGG...
```

This corresponds to the amino acid sequence <SEQ ID 888; ORF112-1>:

```
45 51 MNLISRYIIR QMAVMAVYAL LAFLALYSFF EILYETGNLG KGSYGIWEML
51 GYTALKMPAR AYELIPLAVL IGGLVSLSQL AAGSELTVIK ASGMSTKKLL
101 LILSQFGFIF AIATVALGEW VAPTLSQKAE NIKAAAINGK ISTGNTGLWL
151 KEKNSXINVR EMLPDHTLLG IKIWARNDKN ELAEAVEADS AVLNSDGSWQ
201 LKNIRRSTLG EDKVEVSIAA EENWPISVKR NLMDVLLVKP DQMSVGELTT
251 YIRHLQNNSQ NTRIYAIAWW RKLVYPAAAW VMALVAFAFT PQTTRHGNMG
50 301 LKLFGGICXG LLFHLAGRLF GFTSQL...
```

Computer analysis of this amino acid sequence predicts two transmembrane domains and gave the following results:

# Homology with a predicted ORF from N. meningitidis (strain A)

ORF112 shows 96.4% identity over a 166aa overlap with an ORF (ORF112a) from strain A of N.

meningitidis:

orf112a.pep

orf112a.pep

orf112-1

60

	meningitidis:		
	•	10 00 20 40 50 6	0
_	<b>5440</b>		
5	orf112.pep	MNLISKYIIRQMAVMAVIALEAFEALISFFEILIEIGNIGRGSIGWEMBSTALIGHT	ì
	£1122	MNLISRYIIRQMAVMAVYALLAFLALYSFFEILYETGNLGKGSYGIWEMXGYTALKMXA	Ŕ
	orf112a		0
		10 20 30 10 50	
10		70 80 90 100 110 12	0
10	orf112.pep		
	Offits.beb		1
	orf112a	AYELMPLAVLIGGLVSXSQLAAGSELXVIKASGMSTKKLLLILSQFGFIFAIATVALGE	W
		70 80 90 100 110 12	0
15			
10		130 140 150 160	
	orf112.pep	VAPTLSQKAENIKAAAINGKISTGNTGLWLKEKNSVINVREMLPDH	
	orf112a	VAPTLSQKAENIKAAAINGKISTGNTGLWLKEKNSIINVREMLPDHTLLGIKIWARNDF	
20	•	130 140 150 160 170 18	10
			r D
	orf112a	ELAEAVEADSAVLNSDGSWQLKNIRRSTLGEDKVEVSIAAEEXWPISVKRNLMDVLLVF	10
		190 200 210 220 230 24	10
	The ODE 1120 m	ucleotide sequence <seq 889="" id=""> is:</seq>	
	The OKF 112a III	ucleotide sequence - SEQ ID 669> 15.	
~ ~	_		
25	1	ATGAACCTGA TTTCACGTTA CATCATCCGT CAAATGGCGG TTATGGCGGT TTACGCGCTC CTTGCCTTCC TCGCTTTGTA CAGCTTTTTT GAAATCCTGT	
	51	ACGAAACCGG CAACCTCGGC AAAGGCAGTT ACGGCATATG GGAAATGNTG	
	101 151	GGNTACACCG CCCTCAAAAT GNCCGCCCGC GCCTACGAAC TGATGCCCCT	
	201	CGCCGTCCTT ATCGGCGGAC TGGTCTCTNT CAGCCAGCTT GCCGCCGGCA	
30	251	GCGAACTGAN CGTCATCAAA GCCAGCGGCA TGAGCACCAA AAAGCTGCTG	
30	301	TTGATTCTGT CGCAGTTCGG TTTTATTTTT GCTATTGCCA CCGTCGCGCT	
	351	CGGCGAATGG GTTGCGCCCA CACTGAGCCA AAAAGCCGAA AACATCAAAG	
	401	CCGCGGCCAT CAACGCCAAA ATCAGTACCG GCAATACCGG CCTTTGGCTG	
	451	AAAGAAAAA ACAGCATTAT CAATGTGCGC GAAATGTTGC CCGACCATAC	
35	501	CCTGCTGGGC ATTAAAATCT GGGCCCGCAA CGATAAAAAC GAACTGGCAG	
	551	AGGCAGTGGA AGCCGATTCC GCCGTTTTGA ACAGCGACGG CAGTTGGCAG	
	601	TTGAAAAACA TCCGCCGCAG CACGCTTGGC GAAGACAAAG TCGAGGTCTC	
	651	TATTGCGGCT GAAGAAAANT GGCCGATTTC CGTCAAACGC AACCTGATGG	
	701	ACGTATTGCT CGTCAAACCC GACCAAATGT CCGTCGGCGA ACTGACCACC	
40	751	TACATCCGCC ACCTCCAAAN NNACAGCCAA AACACCCGAA TCTACGCCAT	
	801	CGCATGGTGG CGCAAATTGG TTTACCCCGC CGCAGCCTGG GTGATGGCGC	
	851	TCGTCGCCTT TGCCTTTACC CCGCAAACCA CCCGCCACGG CAATATGGGC TTAAAANTCT TCGGCGGCAT CTGTCTCGGA TTGCTGTTCC ACCTTGCCGG	
	901	NCGGCTCTTC NGGTTTACCA GCCAACTCTA CGGCATCCCG CCCTTCCTCG	
45	951 1001	NCGGCGCACT ACCTACCATA GCCTTCGCCT TGCTCGCCGT TTGGCTGATA	
45	1051	CGCAAACAGG AAAAACGCTA A	
	This encodes a r	protein having the amino acid sequence <seq 890="" id="">:</seq>	
	11115 01100000 0 1	h	
	1	MNLISRYIIR QMAVMAVYAL LAFLALYSFF EILYETGNLG KGSYGIWEMX	
	51	GYTALKMXAR AYELMPLAVL IGGLVSXSQL AAGSELXVIK ASGMSTKKLL	
50	101	THE PARTY OF THE P	
20	151	KEKNSIINVR EMLPDHTLLG IKIWARNDKN ELAEAVEADS AVLNSDGSWQ	
	201	LKNIRRSTLG EDKVEVSIAA EEXWPISVKR NLMDVLLVKP DQMSVGELTT	
	251	YIRHLOXXSO NTRIYAIAWW RKLVYPAAAW VMALVAFAFT PQTTRHGNMG	
	301	LKXFGGICLG LLFHLAGRLF XFTSQLYGIP PFLXGALPTI AFALLAVWLI	
55	351		
-		DDD440.4.1. 0C.00/11 vb. 1.00C1	
	ORF112a and C	ORF112-1 show 96.3% identity in 326 aa overlap:	
		· · · · · · · · · · · · · · · · · · ·	
			22

 ${\tt MNLISRYIIRQMAVMAVYALLAFLALYSFFEILYETGNLGKGSYGIWEMXGYTALKMXAR}$ 

 ${\tt AYELMPLAVLIGGLVSXSQLAAGSELXVIKASGMSTKKLLLILSQFGFIFAIATVALGEW}$ 

PCT/IB98/01665

-485-

	orf112-1	:
5	orf112a.pep	VAPTLSQKAENIKAAAINGKISTGNTGLWLKEKNSIINVREMLPDHTLLGIKIWARNDKN 
10	orf112a.pep	ELAEAVEADSAVLNSDGSWQLKNIRRSTLGEDKVEVSIAAEEXWPISVKRNLMDVLLVKP
	orf112a.pep	DQMSVGELTTYIRHLQXXSQNTRIYAIAWWRKLVYPAAAWVMALVAFAFTPQTTRHGNMG 
15	orf112a.pep	LKXFGGICLGLLFHLAGRLFXFTSQLYGIPPFLXGALPTIAFALLAVWLIRKQEKRX                          LKLFGGICXGLLFHLAGRLFGFTSQL

# 20 Homology with a predicted ORF from N. gonorrhoeae

ORF112 shows 95.8% identity over 166aa overlap with a predicted ORF (ORF112ng) from N. gonorrhoeae:

```
MNLISRYIIRQMAVMAVYALLAFLALYSFFEILYETGNLGKGSYGIWEMLGYTALKMPAR
       orf112.pep
                 MNLISRYIIRQMAVMAVYALLAFLALYSFFEILYETGNLGKGSYGIWEMLGYTALKMPAR 60
25
       orf112ng
                 AYELIPLAVLIGGLVSLSQLAAGSELTVIKASGMSTKKLLLILSQFGFIFAIATVALGEW 120
       orf112.pep
                 AYEIMPLAVLIGGLASLSQLAAGSELAVIKASGMSTKKLLLILSQFGFIFAIAAVALGEW 120
       orf112ng
30
                 VAPTLSQKAENIKAAAINGKISTGNTGLWLKEKNSVINVREMLPDH
       orf112.pep
                 VAPTLSQKAENIKAAAINGKISTGNTGLWLKEKTSIINVRGMLPDHTLLGIKIWARNDKN 180
       orfll2ng
```

The complete length ORF112ng nucleotide sequence <SEQ ID 891> is:

```
1 ATGAACCTGA TTTCACGTTA CATCATCCGC CAAATGGCGG TTATGGCGGT
35
                    TTACGCGCTC CTTGCCTTCC TCGCTTTGTA CAGCTTTTTT GAAATCCTGT
                    ACGAAACCGG CAACCTCGGC AAAGGCAGTT ACGGCATATG GGAAATGCTG
               151 GGCTACACCG CCCTCAAAAT GCCCGCCCGC GCCTACGAAC TCATGCCCCT
               201 CGCCGTCCTC ATCGGCGGAC TGGCCTCTCT CAGCCAGCTT GCCGCCGGCA
                    GCGAACTGGC CGTCATCAAA GCCAGCGGCA TGAGCACCAA AAAGCTGCTG
               251
40
                    TTGATTCTGT CTCAGTTCGG TTTTATTTTT GCTATTGCCG CCGTCGCGCT
               301
               351 CGGCGAATGG GTTGCGCCCA CGCTGAGCCA AAAAGCCGAA AACATCAAag
                401 cCGCCGCCAt taacggCAAA ATCAGCAccg gcAATACCGG CCTTTggcTG
                451 AAAGAAAAA CCAGCATTAT CAATGTGCGC GGAATGTTGC CCGACCATAC
                    GCTTTTGGGC ATCAAAATTT GGGCGCGCAA CGATAAAAAC GAATTGGCAG
45
                501
                551 AGGCAGTGGA AGCCGATTCC GCCGTTTTGA ACAGCGACGG CAGCTGGCAG
                    TTGAAAAACA TCCGCCGCAG CATCATGGGT ACAGACAAAA TCGAAACATC
                601
                    CGCCGCCGC GAAGAAACTT GGCCGATTGC CGTCAGACGC AACCTGATGG
                651
                701 ACGTATTGCT CGTCAAGCCC GACCAAATGT CCGTCGGCGA GCTGACCACC
                751 TACATCCGCC ACCTCCAAAA CAACAGCCAA AACACCCAAA TCTACGCCAT
50
                    CGCATGGTGG CGTAAACTCG TTTACCCCGT CGCCGCATGG GTCATGGCGC
                801
                     TCGTTGCCTT CGCCTTTACG CCGCAAACCA CGCGCCACGG CAATATGGGC
                851
                     TTAAAACTCT TCGGCGGCAT CTGTCTCGGA TTGCTGTTCC ACCTTGCCGG
                901
                951 CAGGCTCTTC GGGTTTACCA GCCAACTCTA CGGCACCCCA CCCTTCCTCG
               1001 CCGGCGCACT GCCTACCATA GCCTTCGCCT TGCTCGCTGT TTGGCTGATA
 55
               1051 CGCAAACAGG AAAAACGTTG A
```

# This encodes a protein having amino acid sequence <SEQ ID 892>:

60 51 101 151 201	LKNIRRSIMG	AYELMPLAVL AIAAVALGEW GMLPDHTLLG TDKIETSAAA	IGGLASLSQL VAPTLSQKAE IKIWARNDKN EETWPIAVRR RKLVYPVAAW	AAGSELAVIK NIKAAAINGK ELAEAVEADS NLMDVLLVKP VMALVAFAFT	ASGMSTAKLL
----------------------------	------------	---------------------------------------------	--------------------------------------------------------------------	--------------------------------------------------------------------	------------

351 RKQEKR\*

ORF112ng and ORF112-1 show 94.2% identity in 326 aa overlap:

		10	20	30	.40	50	60
	orf112ng	MNLISRYIIRQMAV	/MAVYALLAF	LALYSFFEIL	YETGNLGKGS'	<i>'GIWEMLGYTA</i>	LKMPAR
5		11111111111111		1111111111			
	orf112-1	MNLISRYIIRQMAV	MAVYALLAF	LALYSFFEIL:	YETGNLGKGS	GIWEMLGYT?	ALKMPAR
		10	20	30	40	50	60
	.,	70.	80	90	100	110	120
10	orf112ng	AYELMPLAVLIGGI					
10	Official		-			-	
	orf112-1	AYELIPLAVLIGGI					
	OF1112-1	70	80 21201446	90	100	110	120
	•	70	80	90	100	110	120
15		130	140	150	160	· 170	180
	orf112ng	VAPTLSOKAENIKA	AAAINGKIST	GNTGLWLKEK	TSIINVRGML	PDHTLLGIKI	WARNDKN
•	<b>3</b>	1111111111111		THEFT	:1:1111 11	111111111	111111
	orf112-1	VAPTLSOKAENIKA	AAAINGKIST	GNTGLWLKEK			
		130	140	150	160	170	180
20							
		190	200	210	220	230	240
	orf112ng	ELAEAVEADSAVL	NSDGSWQLKN	IRRSIMGTOK	IETSAAAEET	WPIAVRRNLM	DVLLVKP
	-	#	ншіш	11111:1111	:1:1	111:1:111	RHHH
	orf112-1	ELAEAVEADSAVL	NSDGSWOLKN	IRRSTLGEDK	VEVSIAAEEN	WPISVKRNLM	DVLLVKP
25		190	200	210	220	230	240
		250	260	270	280	290	300
	orf112ng	DOMSVGELTTYIRE					
	Offitzing					_	IIIIIIII
30	orf112-1	DOMSVGELTTYIRE					
30	011112-1	250	260	270	280	290	300
		250	260	270	. 200	290	300
		310	320	330	340	350	,
	orfl12ng	LKLFGGICLGLLF	HLAGRLFGF	SQLYGTPPFL	AGALPTIAFA	LLAVWLIRKQ	EKRX
35	•	11111111 1111	11111111111	HÏI		_	
	orf112-1	LKLFGGICXGLLF	HLAGRLFGFT	rsql			
	•	310	320				

This analysis suggests that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

It will be appreciated that the invention has been described by means of example only, and that modifications may be made whilst remaining within the spirit and scope of the invention.

TABLE I – PCR primers

ORF	Primer	Sequence	Restriction sites
ORF 1	Forward Reverse	CGC <u>GGATCCGCTAGC</u> -GGACACACTTATTTCGG CCCG <u>CTCGAG</u> -CCAGCGGTAGCCTAATT	BamHI-NheI XhoI
ORF 2	Forward Reverse	GCGGATCCCATATG-TTTGATTTCGGTTTGGG CCCGCTCGAG-GACGGCATAACGGCG	BamHI-NdeI XhoI
ORF 2-1	Forward Reverse	GCGGATCCCATATG-TTTGATTTCGGTTTGGG CCCGCTCGAG-TGATTTACGGACGCGCA	BamHI-NdeI XhoI
ORF 4	Forward Reverse	GC <u>GGATCCCATATG</u> -TGCGGAGGTCAAAAAGAC CCCG <u>CTCGAG</u> -TTTGGCTGCGCCTTC	BamHI-Ndel XhoI
ORF 5	Forward Forward Reverse	GGAATTCCATATGGCCATGG-TGGAAGGCGCACAACC CGGGATCC-ATGGAAGGCGCACAAC CCCGCTCGAG-GACTGTGCAAAAACGG	NdeI-NcoI BamHI XhoI
ORF 6	Forward Reverse	CGCGGATCCCATATG-ACCCGTCAATCTCTGCA CCCGCTCGAG-TGCGCCGAACACTTTC	BamHI-NdeI XhoI
ORF 7	Forward Reverse	CGCGGATCCGCTAGC-GCGCTGCTTTTTGTTCC CCCGCTCGAG-TTTCAAAATATATTTGCGGA	BamHI-Nhel XhoI
ORF 8	Forward Reverse	GCGGATCCCATATG-GCTCAACTGCTTCGTAC CCCGCTCGAG-AGCAGGCTTTGGCGC	BamHI-Ndel XhoI
ORF 9	Forward Reverse	CGC <u>GGATCCCATATG</u> -CCGAAGGAAGTCGGAAA CCCG <u>CTCGAG</u> -TTTCCGAGGTTTTCGGG	BamHI-Ndel XhoI
ORF 10	Forward Reverse		BamHI-Nde XhoI
ORF 11	Forward Reverse	- CONTROL CONT	BamHI-Nde XhoI
ORF 13	Forward Reverse	ACCOMPANCE ACCOMPANCE ATTACK	BamHI-Nde XhoI
ORF 15	Forward Forward Reverse	CGGGATCC-TGCGGGACACTGACAGG	NdeI-NcoI BamHI XhoI
ORF 17	Forwar	d GGAATTCCATATGGCCATGG -TTGCCGGCCTGTTCG	Ndel-Ncol

	Eammand	CGGGATCC-ATTGCCGGCCTGTTCG	BamHI
	Forward	CCCGCTCGAG-AAGCAGGTTGTACAGC	
	Reverse	CCCGCTCGAG-AAGCAGGTTGTACAGC	XhoI
ORF 18	Forward	  GCGGATCCCATATG-ATTTTGCTGCATTTGGAT	BamHI-NdeI
OKI 18		CCCGCTCGAG-TCTTCCAATTTCTGAAAGC	
·	Reverse	TOTOCAMITICIAMAGO	XhoI
ORF 19	Forward	GGAATTCCATATGGCCATGG -TCGCCAGTGTTTTTACC	NdeI-NcoI
	Forward	CGGGATCC-TTCGCCAGTGTTTTTACCG	BamHI
	Reverse	CCCGCTCGAG-GGTGTTTTTGAAGCTGCC	XhoI
	Keveise		Alloi
ORF 20	Forward	GGAATTCCATATGGCCATGG -TCGGCGCGGGTATG	NdeI-NcoI
	Forward	CGGGATCC-TTCGGCGCGGGTATG	BamHI
	Reverse	CCCGCTCGAG-CGGCGAGCGAGCCA	XhoI
	10000150		
ORF 22	Forward	GGAATTC <u>CATATG</u> G <u>CCATG</u> G-TGATTAAAATCAAAAAAAGGTCT	NdeI-NcoI
	Forward	CGGGATCC-ATGATTAAAATCAAAAAAGGTCTAAACC	BamHI
·	Reverse	CCCGCTCGAG-ATTATGATAGCGGCCC	XhoI
ORF 23	Forward	CGCGGATCCCATATG-GATGTTTCTGTTTCAGAC	BamHI-NdeI
,	Reverse	CCCG <u>CTCGAG</u> -TTTAAACCGATAGGTAAACG	XhoI
ORF 24	Forward	GGAATTCCATATGGCCATGG -TGATGCCGGAAATGGTG	NdeI-NcoI
	Forward	CGGGATCC-ATGATGCCGGAAATGGTG	BamHI
ļ	Reverse	CCCG <u>CTCGAG</u> -TGTCAGCGTGGCGCA	XhoI
ORF 25	Forward	GCGGATCCCATATG-TATCGCAAACTGATTGC	BamHI-NdeI
	Reverse	CCCGCTCGAG-ATCGATGGAATAGCCG	XhoI
		COCCAMOCOCAMA MO. CA COMO A MOCA CMA MAC	
ORF 26	Forward	GCGGATCCCATATG -CAGCTGATCGACTATTC	BamHI-NdeI
	Reverse	CCCGCTCGAG-GACATCGGCGCGTTTT	XhoI
ORF 27	Formuland	GGAATTC <u>CATATG</u> G <u>CCATGG</u> -AGACCTATTCTGTTTA	NdeI-NcoI
Old 27	Forward	CGGGATCC- CAGACCTATTCTGTTTATTTTAATC	BamHI
	Reverse	CCCGCTCGAG-GGGTTCGATTAAATAACCAT	XhoI
	Keveise	<u> </u>	Alloi
ORF 28	Forward	GGAATTCCATATGGCCATGG-ACGGCTGTACGTTGATGT	NdeI-NcoI
014 20	Forward	CGGGATCC-AACGGCTGTACGTTGATG	BamHI
1	Reverse	CCCGCTCGAG-TTTGTCAGAGGAATTCGCG	XhoI
	ICCVCISC		Alloi
ORF 29	Forward	GCGGATCCCATATG -AACGGTTTGGATGCCCG	BamHI-NdeI
	Forward	CGCGGATCCGCTAGC-AACGGTTTGGATGCCCG	BamHI-NheI
	Reverse	CCCGCTCGAG-TTTGTCTAAGTTCCTGATATG	XhoI
	100,0130		1
ORF 32	Forward	CGCGGATCCCATATG-AATACTCCTCCTTTTG	BamHI-NdeI
	Reverse	CCCGCTCGAG-GCGTATTTTTTGATGCTTTG	XhoI
			1200
ORF 33	Forward	GCGGATCCCATATG -ATTGATAGGGATCGTATG	BamHI-NdeI
	Reverse	CCCGCTCGAG-TTGATCTTTCAAACGGCC	XhoI
L	1.070130		12401

			1
ORF 35	Forward	GCGGATCCCATATG-TTCAGAGCTCAGCTT	BamHI-NdeI
OM. 22	LOIMAIG	CGCGGATCCGCTAGC-TTCAGAGCTCAGCTT	BamHI-NheI
	10111010	CCCGCTCGAG-AAACAGCCATTTGAGCGA	XhoI
ORF 37	Forward	GCGGATCCCATATG-GATGACGTATCGGATTTT	BamHI-NdeI
ORF 37	I OI Water I	CCCG <u>CTCGAG</u> -ATAGCCCGCTTTCAGG	XhoI
ORF 58	Forward	CGCGGATCCGCTAGC-TCCGAACGCGAGTGGAT	BamHI-NheI
UKF 56	101114	CCCGCTCGAG-AGCATTGTCCAAGGGGAC	XhoI
ORF 65	Forward	GGAATTCCATATGGCCATGG -TGCTGTATCTGAATCAAG	
OIG 05	101		NdeI-NcoI
	Forward	CGGGATCC-TTGCTGTATCTGAATCAAGG	BamHI
	Reverse	CCCG <u>CTCGAG</u> -CCGCATCGGCAGACA	XhoI
ORF 66	Forward	GCGGATCCCATATG-TACGCATTTACCGCCG	BamHI-NdeI
	Reverse	CCCG <u>CTCGAG</u> -TGGATTTTGCAGAGATGG	XhoI
ORF 72	Forward	CGCGGATCCCATATG- AATGCAGTAAAAATATCTGA	BamHI-NdeI
ORF /2	Reverse	CCCGCTCGAG-GCCTGAGACCTTTGCAA	XhoI
ORF 73	Forward	GCGGATCCCATATG-AGATTTTTCGGTATCGG	BamHI-NdeI
	Reverse	CCCGCTCGAG-TTCATCTTTTTCATGTTCG	XhoI
ORF 75	Forward	GCGGATCCCATATG- TCTGTCTTTCAAACGGC	BamHI-NdeI
	Reverse	CCCGCTCGAG-TTTGTTTTTGCAAGACAG	XhoI
ORF 76	Forward	GATCAGCTAGCCATATG-AAACAGAAAAAAACCGC	NheI-NdeI
	Reverse	CGGGATCC-TTACGGTTTGACACCGTT	BamHI
ORF 79	Forward	CGCGGATCCCATATG-GTTTCCGCCGCCG	BamHI-NdeI
	Reverse	CCCGCTCGAG-GTGCTGATGCGCTTCG	XhoI
ORF 83	Forward	GCGGATCCCATATG-AAAACCCTGCTGCTGC	BamHI-NdeI
J. O. O.	Reverse	CCCGCTCGAG-GCCGCCTTTGCGGC	XhoI
ORF 84	Forward	GCGGATCCCATATG-GCAGAGATCTGTTTG	BamHI-NdeI
	Reverse	CCCGCTCGAG-GTTTGCCGATCCGACCA	XhoI
ORF 85	Forward	CGCGGATCCCATATG- GCGGTTTGGGGCGGA	BamHI-NdeI
	Reverse	anneance maccocccccccc	XhoI
ORF 89	Forward	GGAATTCCATATGGCCATGG-CCATACCTTCTTATCA	NdeI-NcoI
	Forward		BamHI
	Reverse		XhoI
ORF 9	7 Forward	GCGGATCCCATATG-CATCCTGCCAGCGAAC	BamHI-Nde

	Reverse	CCCGCTCGAG-TTCGCCTACGGTTTTTTG	XhoI
	ICOVOISC		
ORF 98	Forward	GC <u>GGATCCCATATG</u> -ACGGTAACTGCGG	BamHI-NdeI
	Reverse	CCCG <u>CTCGAG</u> -TTGTTGTTCGGGCAAATC	Xhol
ORF 100	Forward	GCGGATCCCATATG-TCGGGCATTTACACCG	BamHI-NdeI
	Reverse	CCCG <u>CTCGAG</u> -ACGGGTTTCGGCGGAA	XhoI
ODE 101	D 1	GCGGATCCCATATG-ATTTATCAAAGAAACCTC	BamHI-NdeI
ORF 101	Forward Reverse	CCCGCTCGAG-TTTTCCGCCTTTCAATGT	XhoI
	Venerge		Zillor
ORF 102	Forward	GCGGATCCCATATG-GCAGGGCTGTTTTACC	BamHI-NdeI
014 101	Reverse	CCCGCTCGAG-AAACGGTTTGAACACGAC	XhoI
			·
ORF 103	Forward	GCGGATCCCATATG-AACCACGACATCAC	BamHI-NdeI
	Reverse	CCCGCTCGAG-CAGCCACAGGACGGC	XhoI
ORF 104	Forward	GCGGATCCCATATG-ACGTGGGGAACGC	BamHI-NdeI
OKF 104	Reverse	CCCGCTCGAG-GCGGCGTTTGAACGGC	XhoI
	Reverse		
ORF 105	Forward	GCGGATCCCATATG-ACCAAATTTCAAACCCCTC	BamHI-NdeI
	Reverse	CCCGCTCGAG-TAAACGAATGCCGTCCAG	XhoI
ORF 106	Forward	GCGGATCCCATATG-AGGATAACCGACGGCG	BamHI-NdeI
	Reverse	CCCGCTCGAG-TTTGTTCCCGATGATGTT	XhoI
ORF 109	Forward	GCGGATCCCATATG-GAAGATTTATATATAATACTCG	BamHI-NdeI
OKT 107	Reverse	CCCGCTCGAG-ATCAGCTTCGAACCGAAG	XhoI
	10000		
ORF110	Forward	AAAGAATTC-ATGAGTAAATCCCGTAGATCTCCC	EcoRI
	Reverse	AAACTGCAG-GGAAAACCACATCCGCACTCTGCC	PstI
ORF111	Forward	AAAGAATTC-GCACCGCAAAAGGCAAAAACCGCA	EcoRI
OKMIII	Reverse	AAACTGCAG-TCTGCGCGTTTTCGGGCAGGGTGG	PstI
	Reverse		
ORF113	Forward	AAAGAATTC-ATGAACAAAACCCTCTATCGTGTGATTTTCAACCG	
	Reverse	AAACTGCAG-TTACGAATGCCTGCTTGCTCGACCGTACTG	PstI
ORF115	Forward	AAAGAATTC-TTGCTTGTGCAAACAGAAAAAGACGG	EcoRI
	Reverse	AAAAAAGTCGAC-CTATTTTTTAGGGGCTTTTGCTTTGAAAAGCCTGCC	SalI
0.7745		**************************************	E. DI
ORF119	Forward	AAAGAATTC-TACAACATGTATCAGGAAAACCAATACCG AAACTGCAG-TTATGAAAACAGGCGCAGGGCGGTTTTGCC	EcoRI PstI
	Reverse	ANACISCAS - I INIGAMANCASSCSCASSSCSSI I I ISCC	rsu
ORF120	Forward	AAAGAATTC-GCAAGGCTACCCCAATCCGCCGTG	EcoRI
	Reverse	AAACTGCAG-CGGTTTGGCTGCCTGGCCGTTGAT	PstI
OPE111	F	AAAGAATTC-GCCTTGGTCTGGCTGGTTTTCGC	EcoRI
ORF121	Forward		PstI
	Reverse	1.1.1.01.00 TOTAL	1 or

ORF122	Forward	AAAAAA <u>GTCGAC</u> -ATGTC <i>T</i> TACCG <i>C</i> GCAAGCAGTTC <i>T</i> CC	SalI
	Reverse	AAA <u>CTGCAG</u> -TCAGGAACACAAACGATGACGAATATCCGTATC	PstI
ORF125	Forward	AAA <u>GAATTC</u> -GCGCTGTTTTTTGCGGCGGCGTAT	EcoRI
	Reverse	AAA <u>CTGCAG</u> -CGCCGTTTCAAGACGAAAAAGTCG	PstI
ORF126	Forward	AAA <u>GAATTC</u> -GCGGAAACGGTCGAAG	EcoRI
	Reverse	AAA <u>CTGCAG</u> -TTAATCTTGTCTTCCGATATAC	PstI
ORF127	Forward Reverse	AAAGAATTC-ATGACTGATAATCGGGGGTTTACG AAAAAAGTCGAC-CTTAAGTAACTTGCAGTCCTTATC	EcoRI SalI
ORF128	Forward	AAAGAATTC-ATGCAAGCTGTCCGCTACAGGCC	EcoRI
	Reverse	AAACTGCAG-CTAITGCAATGCGCCGCCGCGGAATGITTGAGCAGGCG	PstI
ORF129	Forward	AAAGAATTC-ATGGATTTTCGTTTTGACATTATTTACGAATACCG	EcoRI
	Reverse	AAACTGCAG-TTATTTTTTGATGAAATTTTTGGGGCGG	PștI
ORF130	Forward Reverse	AAAGAATTC-GCAGTACTTGCCATTCTCGGTGCG AAACTGCAG-CTCCGGATCGTCTGTAAACGCATT	EcoRI PstI
ORF 131	Forward	GC <u>GGATCCCATATG</u> -GAAATTCGGGCAATAAAAT	BamHI-NdeI
	Reverse	CCCG <u>CTCGAG</u> -CCAGCGGACGCGTTC	XhoI
ORF 132	Forward Reverse	GCGGATCCCATATG-AAAGAAGCGGGGTTTG CCCGCTCGAG-CCAATCTGCCAGCCGT	BamHI-NdeI XhoI
ORF 133	Forward	CGC <u>GGATCCCATATG</u> -GAAGATGCAGGGCGCG	BamHI-NdeI
	Reverse	CCCG <u>CTCGAG</u> -AAACTTGTAGCTCATCGT	XhoI
ORF 134	Forward Reverse	GCGGATCCCATATG-TCTGTGCAAGCAGTATTG CCCGCTCGAG-ATCCTGTGCCAATGCG	BamHI-NdeI XhoI
ORF 135	Forward Reverse	GCGGATCCCATATG-CCGTCTGAAAAAGCTTT CCCGCTCGAG-AAATACCGCTGAGGATG	BamHI-NdeI XhoI
ORF 136	Forward	CGC <u>GGATCCGCTAGC</u> -ATGAAGCGGCGTATAGCC	BamHI-Nhel
	Reverse	CCCG <u>CTCGAG</u> -TTCCGAATATTTGGAACTTTT	XhoI
ORF 137	Forward	CGC <u>GGATCCCATATG</u> -GGCACGGCGGGAAATA	BamHI-NdeI
	Reverse	CCCG <u>CTCGAG</u> -ATAACGGTATGCCGCC	XhoI
ORF 138	Forward	GCGGATCCCATATG-TTTCGTTTACAATTCAGGC	BamHI-NdeI
	Reverse	CCCGCTCGAG-CGGCGTTTTATAGCGG	XhoI
ORF 139	Forward Reverse	GCGGATCCCATATG-GCTTTTTTGGCGGTAATG CCCGCTCGAG-TAACGTTTCCGTGCGTTT	BamHI-NdeI XhoI

ORF 140	Forward	GC <u>GGATCCCATATG</u> -TTGCCCACAGGCAGC	BamHI-NdeI
	Reverse	CCCG <u>CTCGAG</u> -GACGATGGCAAACAGC	XhoI
ORF 141	Forward Reverse	GCGGATCCCATATG-CCGTCTGAAGCAGTCT CCCGCTCGAG-ATCTGTTGTTTTTAAAATATT	BamHI-Ndel XhoI
ORF 142	Forward	GC <u>GGATCCCATATG</u> -GATAATTCTGGTAGTGAAG	BamHI-NdeI
	Reverse	CCCG <u>CTCGAG</u> -AAACGTATAGCCTACCT	XhoI
ORF 143	Forward	GC <u>GGATCCCATATG</u> -GATACCGCTTTGAACCT	BamHI-NdeI
	Reverse	CCCG <u>CTCGAG</u> -AATGGCTTCCGCAATATG	XhoI
ORF 144	Forward	GCGGATCCCATATG-ACCTTTTTACAACGTTTGC	BamHI-NdeI
	Reverse	CCCG <u>CTCGAG</u> -AGATTGTTGTTGTTTTTCG	XhoI
ORF 147	Forward	GC <u>GGATCCCATATG</u> -TCTGTCTTTCAAACGGC	BamHI-NdeI
	Reverse	CCCG <u>CTCGAG</u> -TTTGTTTTTGCAAGACAG	XhoI

# NB:

5

- restriction sites are underlined
- for ORFs 110-130, where the ORF itself carries an EcoRI site (eg. ORF122), a SalI site was used in the forward primer instead. Similarly, where the ORF carries a PstI site (eg. ORFs 115 and 127), a SalI site was used in the reverse primer.

TABLE II - Summary of cloning, expression and purification

ORF	PCR/cloning	His-fusion	GST-fusion	Purification
		expression	expression	
orf 1	+	+	+	His-fusion
orf 2	+	+	+	GST-fusion
orf 2.1	+	n.d.	+	GST-fusion
orf 4	+	+	+	His-fusion
orf 5	+	n.d.	+	GST-fusion
orf 6	+	+	+	GST-fusion
orf 7	+	+	+	GST-fusion
orf 8	+	n.d.	n.d.	
orf 9	+	+	+	GST-fusion
orf 10	+	n.d.	n.d.	
orf 11	+	n.d.	n.d.	l
orf 13	+	n.d.	+	GST-fusion
orf 15	+	+	+	GST-fusion
orf 17	+	n.d.	n.d.	
orf 18	+	n.d.	n.d.	
orf 19	+	n.d.	n.d.	
orf 20	+	n.d.	n.d.	
orf 22	+	+	+	GST-fusion
orf 23	+	+	+	His-fusion
orf 24	+	n.d.	n.d.	
orf 25	+	+	+	His-fusion
orf 26	+	n.d.	n.d.	
orf 27	+	+	+	GST-fusion
orf 28	+	+	+	GST-fusion
orf 29	+	n.d.	n.d.	
orf 32	+	+	+	His-fusion
orf 33	+	n.d.	n.d.	
orf 35	+	n.d.	n.d.	
orf 37	+	+	+	GST-fusion
orf 58	+	n.d.	n.d.	
orf 65	+	n.d.	n.d.	
orf 66	+	n.d.	n.d.	
orf 72	+	+	n.d.	His-fusion
orf 73	+	n.d.	+	n.d.
orf 75	+	n.d.	n.d.	
orf 76	+	+	n.d.	His-fusion
orf 79	+	+	n.d.	His-fusion
orf 83	+	n.d.	+	n.d.
orf 84	+	n.d.	n.d.	

orf 85	+	n.d.	+	GST-fusion
orf 89	+	n.d.	+	GST-fusion
orf 97	+	+	+	GST-fusion
orf 98	+	n.d.	n.d.	
orf 100	+	n.d.	n.d.	
orf 101	+	n.d.	n.d.	
orf 102	+	n.d.	n.d.	
orf 103	+	n.d.	n.d.	
orf 104	+	n.d.	n.d.	
orf 105	+	n.d.	n.d.	
orf 106	+	+	+	His-fusion
orf 109	+	n.d.	n.d.	
orf 110	+	n.d.	n.d.	
orf 111	+	+	n.d.	His-fusion
orf 113	+ _	+	n.d.	His-fusion
orf 115	n.d.	n.d.	n.d.	
orf 119	+	+	n.d.	His-fusion
orf 120	+	+	n.d.	His-fusion
orf 121	+	n.d.	n.d.	
orf 122	+	+	n.d.	His-fusion
orf 125	+	+	n.d.	His-fusion
orf 126	+	+	n.d.	His-fusion
orf 127	+	+	n.d.	His-fusion
orf 128	+	n.d.	n.d.	
orf 129	. +	+	n.d.	His-fusion
orf 130	+	n.d.	n.d.	
orf 131	+	+	+	n.d.
orf 132	+	+	+	His-fusion
orf 133	+	n.d.	+	GST-fusion
orf 134	+	n.d.	n.d.	
orf 135	+	n.d.	n.d.	
orf 136	+	n.d.	n.d.	
orf 137	+	n.d.	+	GST-fusion
orf 138	+	n.d.	+	GST-fusion
orf 139	+	n.d.	n.d.	
orf 140	+	n.d.	n.d.	<del>- </del>
orf 141	+	n.d.	n.d.	<del></del>
orf 142	+.	n.d.	n.d.	<del> </del>
orf 143	+	n.d.	n.d.	<del> </del>
orf 144	+	n.d.	+	n.d.
orf 147	+	n.d.	n.d.	

#### **CLAIMS**

30

- 1. A protein comprising an amino acid sequence selected from the group consisting of SEQ IDs 2, 4, 6, and 8.
- 2. A nucleic acid molecule which encodes a protein according to claim 1.
- 5 3. A nucleic acid molecule according to claim 2, comprising a nucleotide sequence selected from the group consisting of SEQ IDs 1, 3, 5, and 7.
- A protein comprising an amino acid sequence selected from the group consisting of SEQ 4. IDs 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 10 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 250, 252, 254, 256, 258, 260, 262, 264, 266, 268, 270, 272, 274, 276, 278, 280, 282, 284, 286, 288, 290, 292, 294, 296, 298, 300, 302, 304, 306, 308, 310, 312, 314, 316, 318, 320, 322, 324, 326, 328, 330, 332, 334, 336, 338, 340, 342, 15 344, 346, 348, 350, 352, 354, 356, 358, 360, 362, 364, 366, 368, 370, 372, 374, 376, 378, 380, 382, 384, 386, 388, 390, 392, 394, 396, 398, 400, 402, 404, 406, 408, 410, 412, 414, 416, 418, 420, 422, 424, 426, 428, 430, 432, 434, 436, 438, 440, 442, 444, 446, 448, 450, 452, 454, 456, 458, 460, 462, 464, 466, 468, 470, 472, 474, 476, 478, 480, 482, 484, 486, 488, 490, 492, 494, 496, 498, 500, 502, 504, 506, 508, 510, 512, 514, 516, 518, 520, 522, 524, 526, 528, 530, 532, 534, 536, 538, 540, 542, 20 544, 546, 548, 550, 552, 554, 556, 558, 560, 562, 564, 566, 568, 570, 572, 574, 576, 578, 580, 582, 584, 586, 588, 590, 592, 594, 596, 598, 600, 602, 604, 606, 608, 610, 612, 614, 616, 618, 620, 622, 624, 626, 628, 630, 632, 634, 636, 638, 640, 642, 644, 646, 648, 650, 652, 654, 656, 658, 660, 662, 664, 666, 668, 670, 672, 674, 676, 678, 680, 682, 684, 686, 688, 690, 692, 694, 696, 698, 700, 702, 704, 706, 708, 710, 712, 714, 716, 718, 720, 722, 724, 726, 728, 730, 732, 734, 736, 738, 740, 742, 25 744, 746, 748, 750, 752, 754, 756, 758, 760, 762, 764, 766, 768, 770, 772, 774, 776, 778, 780, 782, 784, 786, 788, 790, 792, 794, 796, 798, 800, 802, 804, 806, 808, 810, 812, 814, 816, 818, 820, 822, 824, 826, 828, 830, 832, 834, 836, 838, 840, 842, 844, 846, 848, 850, 852, 854, 856, 858, 860, 862, 864, 866, 868, 870, 872, 874, 876, 878, 880, 882, 884, 886, 888, 890, & 892...
  - 5. A protein having 50% or greater sequence identity to a protein according to claim 4.

- A protein comprising a fragment of an amino acid sequence selected from the group 6. consisting of SEQ IDs 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 250, 252, 254, 256, 258, 260, 262, 264, 266, 268, 270, 272, 274, 276, 278, 280, 282, 284, 286, 288, 290, 292, 294, 296, 298, 300, 302, 304, 306, 308, 310, 312, 314, 316, 318, 320, 322, 324, 326, 328, 330, 332, 334, 336, 338, 340, 342, 344, 346, 348, 350, 352, 354, 356, 358, 360, 362, 364, 366, 368, 370, 372, 374, 10 376, 378, 380, 382, 384, 386, 388, 390, 392, 394, 396, 398, 400, 402, 404, 406, 408, 410, 412, 414, 416, 418, 420, 422, 424, 426, 428, 430, 432, 434, 436, 438, 440, 442, 444, 446, 448, 450, 452, 454, 456, 458, 460, 462, 464, 466, 468, 470, 472, 474, 476, 478, 480, 482, 484, 486, 488, 490, 492, 494, 496, 498, 500, 502, 504, 506, 508, 510, 512, 514, 516, 518, 520, 522, 524, 526, 528, 530, 532, 534, 536, 538, 540, 542, 544, 546, 548, 550, 552, 554, 556, 558, 560, 562, 564, 566, 568, 570, 572, 574, 15 576, 578, 580, 582, 584, 586, 588, 590, 592, 594, 596, 598, 600, 602, 604, 606, 608, 610, 612, 614, 616, 618, 620, 622, 624, 626, 628, 630, 632, 634, 636, 638, 640, 642, 644, 646, 648, 650, 652, 654, 656, 658, 660, 662, 664, 666, 668, 670, 672, 674, 676, 678, 680, 682, 684, 686, 688, 690, 692, 694, 696, 698, 700, 702, 704, 706, 708, 710, 712, 714, 716, 718, 720, 722, 724, 726, 728, 730, 732, 734, 736, 738, 740, 742, 744, 746, 748, 750, 752, 754, 756, 758, 760, 762, 764, 766, 768, 770, 772, 774, 20 776, 778, 780, 782, 784, 786, 788, 790, 792, 794, 796, 798, 800, 802, 804, 806, 808, 810, 812, 814, 816, 818, 820, 822, 824, 826, 828, 830, 832, 834, 836, 838, 840, 842, 844, 846, 848, 850, 852, 854, 856, 858, 860, 862, 864, 866, 868, 870, 872, 874, 876, 878, 880, 882, 884, 886, 888, 890, & 892...
  - 7. An antibody which binds to a protein according to any one of claims 4 to 6.
- 25 8. A nucleic acid molecule which encodes a protein according to any one of claims 4 to 6.
- 9. A nucleic acid molecule according to claim 8, comprising a nucleotide sequence selected from the group consisting of SEQ IDs 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 135, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253, 255, 257, 259, 261, 263, 265, 267, 269, 271, 273, 275, 277, 279, 281, 283, 285, 287, 289,

291, 293, 295, 297, 299, 301, 303, 305, 307, 309, 311, 313, 315, 317, 319, 321, 323, 325, 327, 329, 331, 333, 335, 337, 339, 341, 343, 345, 347, 349, 351, 353, 355, 357, 359, 361, 363, 365, 367, 369, 371, 373, 375, 377, 379, 381, 383, 385, 387, 389, 391, 393, 395, 397, 399, 401, 403, 405, 407, 409, 411, 413, 415, 417, 419, 421, 423, 425, 427, 429, 431, 433, 435, 437, 439, 441, 443, 445, 447, 449, 451, 453, 455, 457, 459, 461, 463, 465, 467, 469, 471, 473, 475, 477, 479, 481, 483, 485, 487, 489, 5 491, 493, 495, 497, 499, 501, 503, 505, 507, 509, 511, 513, 515, 517, 519, 521, 523, 525, 527, 529, 531, 533, 535, 537, 539, 541, 543, 545, 547, 549, 551, 553, 555, 557, 559, 561, 563, 565, 567, 569, 571, 573, 575, 577, 579, 581, 583, 585, 587, 589, 591, 593, 595, 597, 599, 601, 603, 605, 607, 609, 611, 613, 615, 617, 619, 621, 623, 625, 627, 629, 631, 633, 635, 637, 639, 641, 643, 645, 647, 649, 651, 653, 655, 657, 659, 661, 663, 665, 667, 669, 671, 673, 675, 677, 679, 681, 683, 685, 687, 689, 10 691, 693, 695, 697, 699, 701, 703, 705, 707, 709, 711, 713, 715, 717, 719, 721, 723, 725, 727, 729, 731, 733, 735, 737, 739, 741, 743, 745, 747, 749, 751, 753, 755, 757, 759, 761, 763, 765, 767, 769, 771, 773, 775, 777, 779, 781, 783, 785, 787, 789, 791, 793, 795, 797, 799, 801, 803, 805, 807, 809, 811, 813, 815, 817, 819, 821, 823, 825, 827, 829, 831, 833, 835, 837, 839, 841, 843, 845, 847, 849, 851, 853, 855, 857, 859, 861, 863, 865, 867, 869, 871, 873, 875, 877, 879, 881, 883, 885, 887, 889, 15 & 891..

A nucleic acid molecule comprising a fragment of a nucleotide sequence selected from the 10. group consisting of SEQ IDs 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 20 135, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253, 295, 297, 299, 301, 303, 305, 307, 309, 311, 313, 315, 317, 319, 321, 323, 325, 327, 329, 331, 333, 25 335, 337, 339, 341, 343, 345, 347, 349, 351, 353, 355, 357, 359, 361, 363, 365, 367, 369, 371, 373, 375, 377, 379, 381, 383, 385, 387, 389, 391, 393, 395, 397, 399, 401, 403, 405, 407, 409, 411, 413, 415, 417, 419, 421, 423, 425, 427, 429, 431, 433, 435, 437, 439, 441, 443, 445, 447, 449, 451, 453, 455, 457, 459, 461, 463, 465, 467, 469, 471, 473, 475, 477, 479, 481, 483, 485, 487, 489, 491, 493, 495, 497, 499, 501, 503, 505, 507, 509, 511, 513, 515, 517, 519, 521, 523, 525, 527, 529, 531, 533, 30 535, 537, 539, 541, 543, 545, 547, 549, 551, 553, 555, 557, 559, 561, 563, 565, 567, 569, 571, 573, 575, 577, 579, 581, 583, 585, 587, 589, 591, 593, 595, 597, 599, 601, 603, 605, 607, 609, 611, 613, 615, 617, 619, 621, 623, 625, 627, 629, 631, 633, 635, 637, 639, 641, 643, 645, 647, 649, 651, 653, 655, 657, 659, 661, 663, 665, 667, 669, 671, 673, 675, 677, 679, 681, 683, 685, 687, 689, 691, 693, 35

- 735, 737, 739, 741, 743, 745, 747, 749, 751, 753, 755, 757, 759, 761, 763, 765, 767, 769, 771, 773, 775, 777, 779, 781, 783, 785, 787, 789, 791, 793, 795, 797, 799, 801, 803, 805, 807, 809, 811, 813, 815, 817, 819, 821, 823, 825, 827, 829, 831, 833, 835, 837, 839, 841, 843, 845, 847, 849, 851, 853, 855, 857, 859, 861, 863, 865, 867, 869, 871, 873, 875, 877, 879, 881, 883, 885, 887, 889, & 891...
- 5 11. A nucleic acid molecule comprising a nucleotide sequence complementary to a nucleic acid molecule according to any one of claims 8 to 10.
  - 12. A nucleic acid molecule comprising a nucleotide sequences having 50% or greater sequence identity to a nucleic acid molecule according to any one of claims 8-11.
- 13. A nucleic acid molecule which can hybridise to a nucleic acid molecule according to any one of claims 8-12 under high stringency conditions.
  - 14. A composition comprising a protein, a nucleic acid molecule, or an antibody according to any preceding claim.
  - 15. A composition according to claim 14 being a vaccine composition or a diagnostic composition.
- 15 16. A composition according to claim 14 or claim 15 for use as a pharmaceutical.
  - 17. The use of a composition according to claim 14 in the manufacture of a medicament for the treatment or prevention of infection due to Neisserial bacteria.

1/24

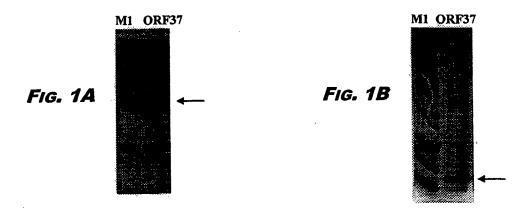
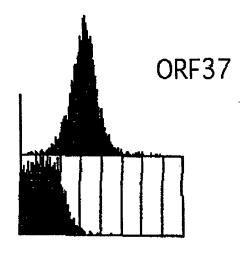


Fig. 1C



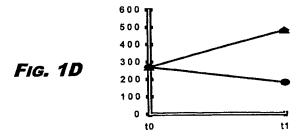
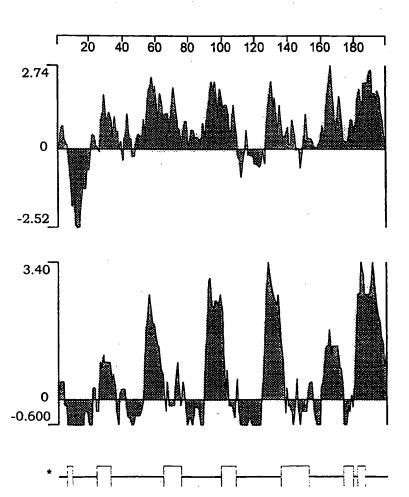
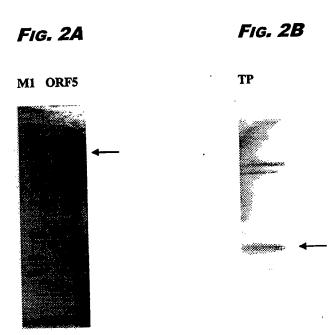


Fig 1E



PCT/IB98/01665

3/24



4/24

FIG. 3A

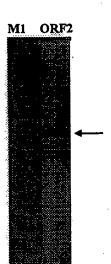


Fig. 3B



FIG. 3C

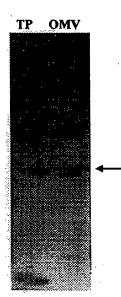
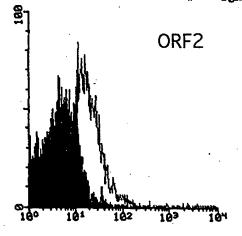


Fig. 3D



PCT/IB98/01665

5/24

FIG. 4A

M1 ORF15

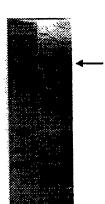


FIG. 4B

M2 ORF15

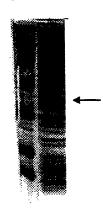
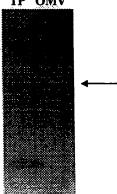


Fig 4C

TP OMV



6/24

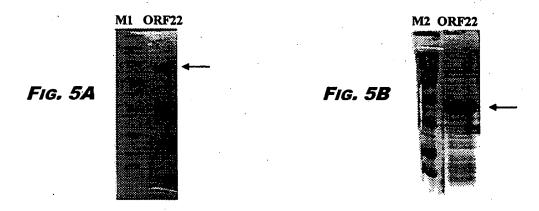
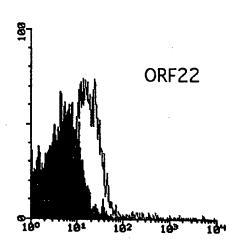
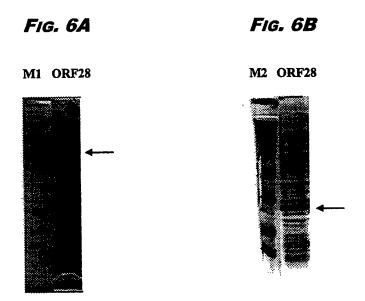


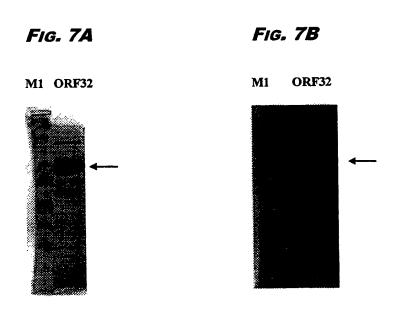
Fig. 5C



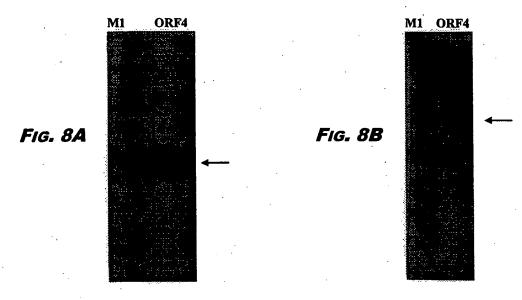
PCT/IB98/01665

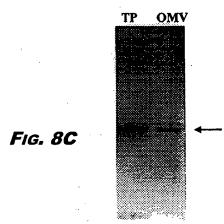
7/24 **FIGURE 6** 





8/24





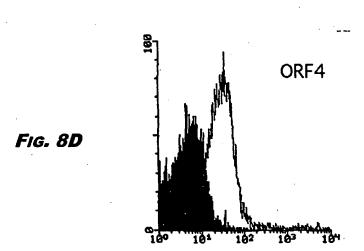


FIG. 8E

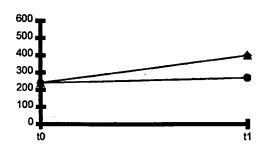
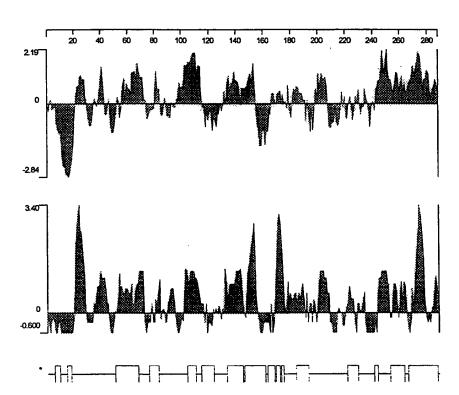
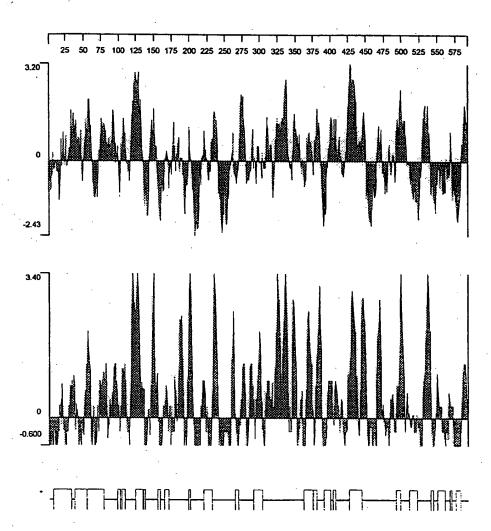


Fig. 8F



10/24



11/24

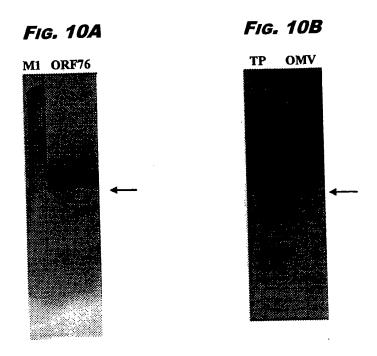
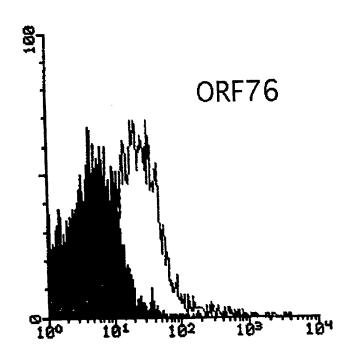
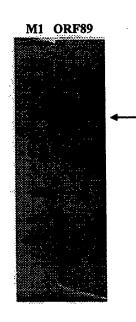


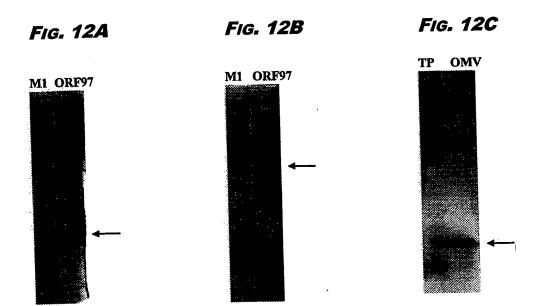
FIG. 10C



12/24



13/24



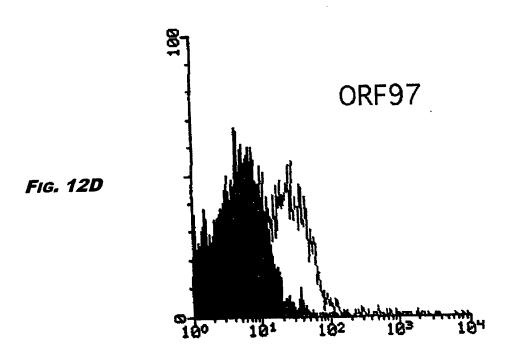
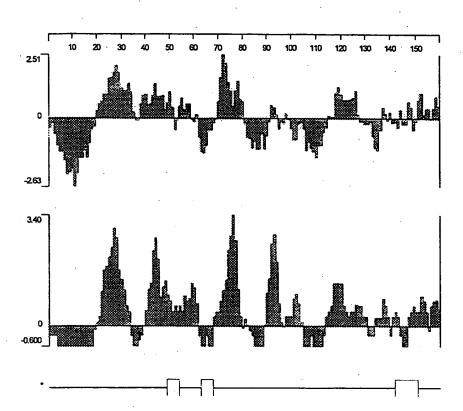
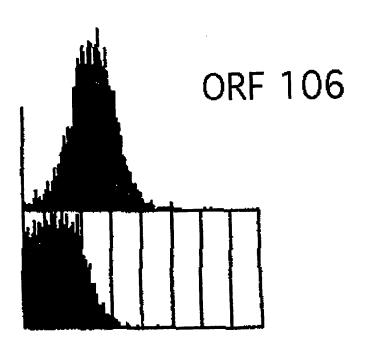


Fig. 12E



15/24

FIG. 13C



16/24

FIGURE 14

Fig. 14A

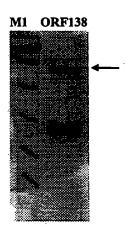
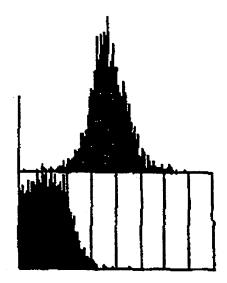


FIG. 14B



17/24

FIG. 15A

M1 ORF23



FIG. 15B

M2 ORF23



FIG 15C

TP OMV

18/24

Fig. 16A



FIG. 16B

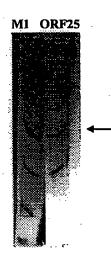


FIG. 16C



Fig. 16D

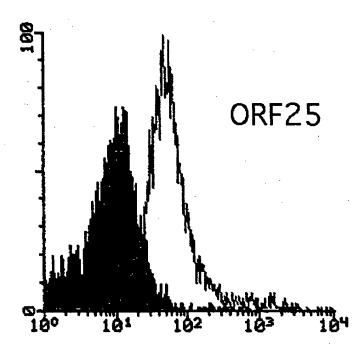


FIG. 16E

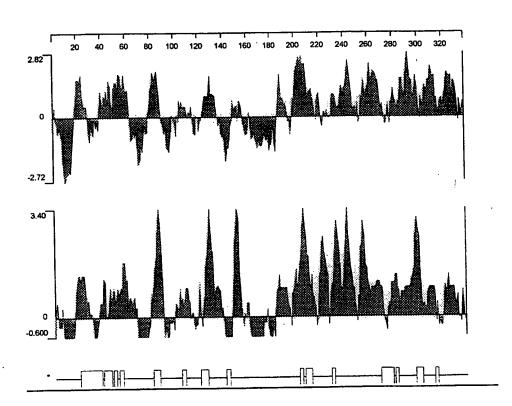


FIG. 17A

M1 ORF27



FIG. 17B

M2 ORF27

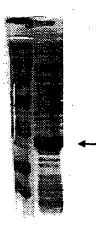
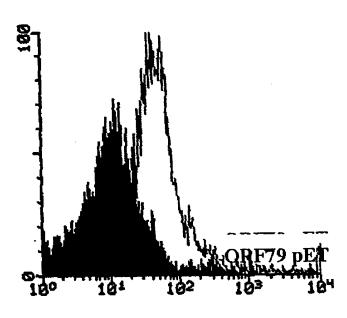


FIGURE 18

FIG. 18A



FIG. 18B



22/24 **FIGURE 19** 

FIG. 19A

MI ORF85

TP OMV

Fig. 19C

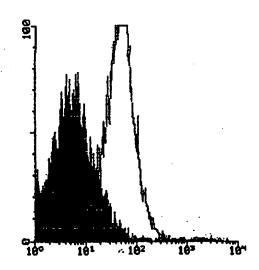
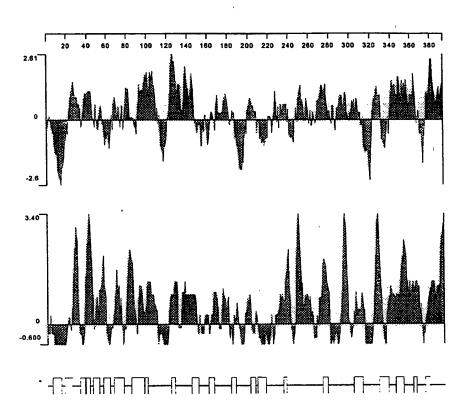


FIG 19D



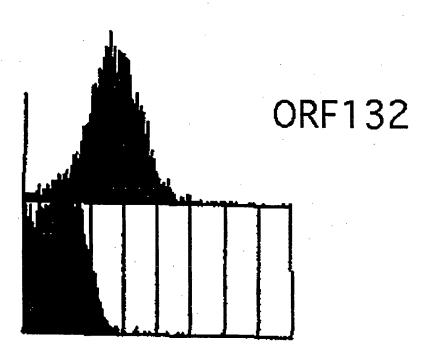
24/24 **FIGURE 20** 

FIG. 20B

M1 ORF132

M2 ORF132

FIG. 20C



#### **PCT**

## WORLD INTELLECTUAL PROPERTY ORGANIZATION International Bureau



#### INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Paten			(11) International Publication Number: WO 99/2	4578
	14/22, 16/12, G01N 33/53, 2N 15/31, C12 R 1:36)	<b>A3</b>	(43) International Publication Date: 20 May 1999 (20	.05.99)
(21) International Appli	cation Number: PCT/IB	98/016	(74) Agent: HALLYBONE, Huw, George; Carpmaels & Ra 43 Bloomsbury Square, London WC1A 2RA (GB).	nsford,
(22) International Filing	<b>Date:</b> 9 October 1998 (	09.10.9		
(30) Priority Data: 9723516.2 9724190.5 9724386.9 9725158.1 9726147.3 9800759.4 9819016.8  (71) Applicant (for all S.P.A. [IT/IT]; N	6 November 1997 (06.11.97) 14 November 1997 (14.11.9) 18 November 1997 (18.11.9) 27 November 1997 (27.11.5) 10 December 1997 (10.12.9) 14 January 1998 (14.01.98) 1 September 1998 (01.09.98)  designated States except US): Via Fiorentina, 1, I–53100 Siena (	07) (0 07) (0 07) (0 07) (0 8) (	KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MI MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YO ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UC Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, T European patent (AT, BE, CH, CY, DE, DK, ES, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI pate BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, N	B, GD, P, KR, K, MN, SI, SK, U, ZW, G, ZW), TJ, TM), FI, FR, ent (BF,
(72) Inventors; and (75) Inventors/Applicants (for US only): MASIGNANI, Vega [IT/IT]; Via Pantaneto, 105, I-53100 Siena (IT). RAP- PUOLI, Rino [IT/IT]; Via delle Rocche, 1, Vagliagli, I-53019 Castelnuovo Berardenga (IT). PIZZA, Mariagrazia [IT/IT]; Strada di Montalbuccio, 160, I-53100 Siena (IT). SCARLATO, Vincenzo [IT/IT]; Via Firenze, 3/37, I-53134 Colle Val d'Elsa (IT). GRANDI, Guido [IT/IT]; 9° Strada, 4, I-20090 Segrate (IT).			(88) Date of publication of the international search report:  2 March 2000 (0)	i2.03.00

(54) Title: NEISSERIAL ANTIGENS

#### (57) Abstract

The invention provides proteins from *Neisseria meningitidis* (strains A and B) and from *Neisseria gonorrhoerae* including amino acid sequences, the corresponding nucleotide sequences, expression data, and serological data. The proteins are useful antigens for vaccines, immunogenic compositions, and/or diagnostics.

#### FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
ΑU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav	TM	Turkmenistan
BF	Burkina Faso	GR	Greece		Republic of Macedonia	TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IÈ	Ireland	MN	Mongolia	UA.	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	· IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	zw	Zimbabwe
Ci	Côte d'Ivoire	KP	Democratic People's	NZ	New Zealand		
CM	Cameroon		Republic of Korea	PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		•
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

PC1/IB 98/01665

a. classification of subject matter IPC 6 C12N15/31 C07K14/22 C07K16/12 G01N33/53 A61K39/095 C12N15/31 //(C12N15/31,C12R1:36) According to International Patent Classification (IPC) or to both national classification and IPC 8. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) C12N C07K G01N A61K IPC 6 Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Electronic data base consulted during the international search (name of data base and, where practical, search terms used) C. DOCUMENTS CONSIDERED TO BE RELEVANT Relevant to claim No. Citation of document, with indication, where appropriate, of the relevant passages Category ' WO 99 55873 A (SMITHKLINE BEECHAM 5-8. Ε 10-16 BIOLOGICALS S.A..) 4 November 1999 (1999-11-04) SEQ ID NOS: 1, 2, 3 and 4 page 82 -page 96 claims 3-5,10,11,17-21 -/--Further documents are listed in the continuation of box C. Х Patent family members are listed in annex. X Special categories of cited documents: "I" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier document but published on or after the international "X" document of particular relevance; the claimed invention filing date cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such docu-\*O\* document referring to an oral disclosure, use, exhibition or ments, such combination being obvious to a person skilled document published prior to the international filing date but \*&\* document member of the same patent family later than the priority date claimed Date of the actual completion of the international search Date of mailing of the international search report OB December 1999 (08.12.99) 1 December 1999 Authorized officer Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentiaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo ni, Fuchs, U Fax: (+31-70) 340-3016

6

International Application No

	· .	PC1, 1B 98	3/01665			
C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT						
Category °	Citation of document, with indication, where appropriate, of the relevant passages		Relevant to claim No.			
A	ROKBI, B. ET AL.: "Evaluation of Recombinant Transferrin-Binding Protein B Variants from Neisseria meningitidis for Their Ability To Induce Cross-Reactive and Bactericidal Antibodies against a Genetically Diverse Collection of		1-17			
	Serogroup B Strains" INFECTION AND IMMUNITY, vol. 65, no. 1, January 1997 (1997-01), pages 55-63, XP002086937 abstract	·				
	page 55, column 1, line 1 -page 56, column 2, line 78 page 57, column 1, line 31 -page 61, column 2, line 63 page 59; figure 3; table 2					
<b>A</b>	WO 96 12020 A (OREGON HEALTH SCIENCES UNIVERITY) 25 April 1996 (1996-04-25) abstract page 32 -page 38; examples 2-4		1-17			
A .	page 42 -page 44; example 9 page 45; table 3 WO 96 31618 A (THE UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL)		1-17			
	10 October 1996 (1996-10-10) abstract page 19, line 21 -page 20, line 13 page 20, line 22 -page 21, line 11 page 22, line 23 -page 24, line 16					

In. ational application No. PCT/IB 98/01665

MIEUMATIONAL	
Box I Observations where certain claims we	re found unsearchable (Continuation of item 1 of first sheet)
	ned in respect of certain claims under Article 17(2)(a) for the following reasons:
Claims Nos.:     because they relate to subject matter not requ	uired to be searched by this Authority, namely:
Claims Nos.:  because they relate to parts of the Internation an extent that no meaningful International Se	nal Application that do not comply with the prescribed requirements to such earch can be carried out, specifically:
ł	not drafted in accordance with the second and third sentences of Rule 6.4(a).
Observations where unity of invention	on is lacking (Continuation of item 2 of first sheet)
	inventions in this international application, as follows:
see additional sheet	
searchable dialms.	timely paid by the applicant, this International Search Report covers all
As all searchable claims could be searche of any additional fee.	d without effort justifying an additional fee, this Authority did not invite payment
COVERS ONLY (LIOSE CIETIES IOI WITHOUT	earch fees were timely paid by the applicant, this International Search Report ere paid, specifically claims Nos.:
((1-3) completely) and (	(4-17) partially)
No required additional search fees were restricted to the invention first mentioned	timely paid by the applicant. Consequently, this International Search Report is in the claims; it is covered by claims Nos.:
Remark on Protest	The additional search fees were accompanied by the applicant's protest.  X  No protest accompanied the payment of additional search fees.

#### FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

Invention 1.Claims: ((1-3) completely) and ((4-17) partially)

A protein comprising an amino acid sequence selected from the group consisting of SEQ ID NOS: 2,4,6 and 8 or fragments thereof; a protein having 50% or greater sequence identity to said protein; an antibody binding said protein; a nucleic acid encoding said protein; a nucleic acid comprising a sequence selected from the group consisting of SEQ ID NOS: 1,3,5 and 7 or fragments thereof; a composition comprising said protein, said nucleic acid or said antibody; the use of said composition;

Invention 2 to 104. Claims (4-17) partially

Idem as subject 1 but limited to the ORFs corresponding to examples 2-104 characterized by SEQ ID NOS: 9-892. (Invention 2 is limited to SEQ ID NOS: 9-10; Invention 3 is limited to SEQ ID NOS: 11-18; Invention 4 is limited to SEQ ID NOS: 19-28; .....; Invention 104 is limited to SEQ ID NOS: 885-892).

In view of additional search fees paid, Inventions 5, 26, 55, 77 and 91 have been further searched.

I nation on patent family members

PC1, 1B 98/01665

Patent document cited in search report		Publication date	Patent family member(s)	Publication date
WO 9955873	A	04-11-1999	NONE	
WO 9612020	A	25-04-1996	US 5698438 A AU 705509 B AU 4007395 A CA 2203116 A EP 0793720 A FI 971634 A HU 77048 A JP 10508469 T NO 971768 A	16-12-1997 27-05-1999 06-05-1996 25-04-1996 10-09-1997 16-06-1997 02-03-1998 25-08-1998 03-06-1997
WO 9631618	Α	10-10-1996	AU 5537096 A CA 2217522 A EP 0830456 A JP 11503322 T	23-10-1996 10-10-1996 25-03-1998 26-03-1999